



2. WYDANIE
UAKTUALNIONE I
ROZSZERZONE

DATA DRIVEN CONSTRUCTION

PORUSZANIE SIĘ PO ERZE DANYCH
W BRANŻY BUDOWLANEJ

Z PRZYPADKAMI
UŻYCIA AI I LLM

Artem Boiko



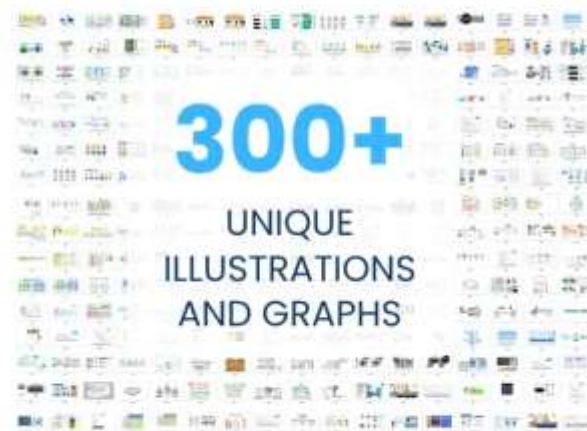
100+

KEY DATA-
RELATED
TOPICS



50+

PRACTICAL
PROBLEMS
SOLVED



300+

UNIQUE
ILLUSTRATIONS
AND GRAPHS

DATA-DRIVEN CONSTRUCTION

Nawigacja w erze danych
w branży budowlanej

Wydanie drugie, poprawione i uzupełnione

ARTEM BOIKO

“

"Boiko jest Jamesem Carville'em IT - w często cytowanym przez tego ostatniego "It's the economy, stupid", tylko jedno słowo musi zostać zamienione na tę słynną książkę. "To dane, głupce". (Aby odnaleźć się we wszechświecie danych, powiedzenie starożytnych Rzymian z czasów greckich jest nadal aktualne: "Navigare necesse est". Autor prowadzi swoich czytelników przez wszystkie głębiny i mielizny oceanu danych z pewną ręką i niezachwianym kompasem, nie wspominając o kompleksowym podejściu historycznym i, co nie mniej ważne, bardzo oryginalnej grafice i dobrym poczuciu humoru, które jest widoczne nie tylko na drugi rzut oka. Międzynarodowa reakcja na książkę Boiko waha się od euforycznej aprobaty po raczej zgryźliwy sceptyczym, co dobrze przysłużyło się drugiemu niemieckiemu wydaniu książki. Boiko jest oryginalnym i niedogmatycznym myślicielem danych. Przedstawia czytelnikowi ekscytujące spostrzeżenia i zawsze odważne, a nawet prowokacyjne tezy, które inspirują do dalszych przemyśleń. Doskonałe lekarstwo na niemiecką chorobę utajonego konsensualizmu. Nawiąsem mówiąc, powyższe łacińskie przysłowie ma dopełnienie: "vivere non est necesse". Nie ma ono zastosowania do podejścia Boiko do świata danych - dane żyją, a ich życie jest konieczne, żeby nie powiedzieć kluczowe".

- Dr Burkhard Talebitari, niezależny redaktor - w tym dla czasopisma: BIM, wydawanego corocznie przez Ernst & Sohn od 2013 roku.

"Książka Artema Boiko jest kamieniem milowym dla demokratyzacji cyfryzacji w branży budowlanej - i prawdziwym przełomem dla małych i średnich przedsiębiorstw (MŚP). Szczególnie przełomowa: dzięki zastosowaniu nowoczesnych narzędzi open-source o niskim kodzie i bez kodu, firmy mogą już teraz skutecznie integrować dane ze swoimi procesami biznesowymi i analizować je z zyskiem - bez dogębnnej wiedzy programistycznej. Sprawia to, że kosztowne korzystanie z uciążliwych komercyjnych pakietów oprogramowania staje się zbędne. Ta książka jest wezwaniem do działania! Jest to cenny przewodnik dla każdego, kto nie tylko chce zrozumieć cyfrową transformację w branży budowlanej, ale także chce ją aktywnie kształtać - pragmatycznie, skutecznie i przyszłościowo. Nadszedł czas, aby wspólnie dzielić się tą wiedzą i trwałe zwiększać produktywność branży budowlanej".

- **Dr Michael Max Buehler**, profesor zarządzania budownictwem w HTWG Konstanz, współwłaściciel GemeinWerk Ventures i niezależny dyrektor DevvStream.

"Książka DataDrivenConstruction jest jednym z pierwszych kroków poza granice zwykłego świata budowniczych, z ich złożonymi systemami projektowania i zarządzania, gdy, jak się wydaje, złożoność i nasycenie danych nie daje nawet szansy na radykalne uproszczenie i zwiększenie przejrzystości pracy z danymi budowlanymi. W swojej książce Artem pokazuje prostym językiem, jakie możliwości otwierają przed nami nowoczesne technologie pracy z danymi i dosłownie podaje konkretne kroki, które można od razu zastosować w swojej pracy. Zachęcam wszystkich, którzy chcą zrozumieć, dokąd zmierzają systemy automatyzacji w branży budowlanej, do uważnego przestudiowania tej książki, aby uświadomić sobie, że rewolucja danych w budownictwie już puka do naszych drzwi. Teraz jest to interesujące tylko dla maniaków, ale za kilka lat, podobnie jak BIM, takie podejście i oprogramowanie będą wszelkie!"

- **Ihor Rogachew**, szef Centrum Kompetencyjnego IMT, BIM & Digital Transformation w RGD i założyciel InfraBIM.Pro.

"Gorąco polecam książkę DataDrivenConstruction, która dotyczy, jak mówi tytuł, podejścia do zarządzania informacjami opartego na danych dla AECO. Obecnie korzystam z niej, aby zainicjować szereg dyskusji z różnymi grupami. Uważam, że jest to bardzo przystępne odniesienie. Oprócz dokładnego przeglądu kontekstu historycznego narzędzi w AECO, danych i wprowadzenia kilku kluczowych technologii, książka zawiera szereg bardzo przydatnych diagramów, które przedstawiają zakres źródeł danych i artefaktów użytkownika końcowego wraz z przykładowymi przepływami pracy. Uderza mnie, że są to rodzaje diagramów, których potrzebujemy więcej przy opracowywaniu i monitorowaniu strategii informacyjnych i przyczyniają się do BEP - definiując ogólny model danych przedsiębiorstwa, na który można nałożyć granicę dla PIM i AIM".

- **Paul Ransley**, główny konsultant w Acmena i inżynier ds. integracji systemów w Transport for London.

"Jeśli "dane są nową ropą naftową", musimy nauczyć się je definiować, znajdować, wydobywać i udoskonalać, aby uczynić je wartościowymi. Uważam, że książka DataDrivenConstruction jest bardzo pouczająca i wnikienna. Książka zapewnia przydatne tło historyczne i wyjaśnia pracę z danymi prostym językiem. Dla tych, którzy są zainteresowani transformacją cyfrową, daje ona dobre zrozumienie danych - jak działają, jak są ustrukturyzowane i jak można je wykorzystać".

- **Ralph Montague**, dyrektor w ArcDox, dyrektor BIM Coordinators Summit i przewodniczący BIM National Mirror Committee w National Standards Authority of Ireland.

"Jak podkreślono w książce, informacje są kluczowym zasobem dla sektora budowlanego, a posiadanie ich w przystępnych formatach znacznie ułatwia podejmowanie trafnych decyzji i przyspiesza harmonogramy projektów. Książka oferuje neutralne i skuteczne podejście do uzyskiwania dostępu i korzystania z tego źródła przy podejmowaniu decyzji. Metodologia przedstawiona w książce wykorzystuje współczesne podejście, które łączy programowanie oparte na sztucznej inteligencji z dostępnymi narzędziami typu open source. Wykorzystując moc sztucznej inteligencji i wykorzystując oprogramowanie open-source, metodologia ta ma na celu zwiększenie automatyzacji, optymalizację procesów oraz promowanie dostępności i współpracy w tej dziedzinie. Język książki jest przejrzysty i łatwy do zrozumienia".

- **Dr Salih Ofluoğlu**, dziekan Wydziału Sztuk Pięknych i Architektury na Uniwersytecie Antalya Bilim oraz organizator Euroazjatyckiego Forum BIM.

"Wszystko, co mogę powiedzieć, to WOW! Sposób, w jaki włączyłeś historię, LLM, grafikę i ogólną łatwość zrozumienia twoich punktów jest naprawdę niezwykły. Plynność książki jest niesamowita. Jest tak wiele genialnych aspektów tej książki; to naprawdę zmienia zasady gry. To świetne źródło informacji i pochwalam cię za wysiłek i pasję, które w nią włożyłeś. Gratuluję stworzenia tak niezwykłego dzieła. Mógłbym kontynuować, ale wystarczy powiedzieć, że jestem pod ogromnym wrażeniem!"

- **Natasha Prinsloo**, Digital Practice Lead w energylab_.

"Dla każdego w branży budowlanej, od nowicjuszy po doświadczonych profesjonalistów, ta książka jest przełomem! Nie jest to typowa zakurzona lektura - jest pełna spostrzeżeń, strategii i odrębny humor, aby utrzymać zaangażowanie. Od starożytnych metod rejestrowania danych po najnowocześniejsze technologie cyfrowe, obejmuje ewolucję wykorzystania danych w budownictwie. To jak podróz wehikułem czasu przez ewolucję danych budowlanych. Niezależnie od tego, czy jesteś architektem, inżynierem, kierownikiem projektu czy analitykiem danych, ten kompleksowy przewodnik zrewolucjonizuje sposób, w jaki podejmujesz do projektów. Przygotuj się na optymalizację procesów, usprawnienie procesu decyzyjnego i zarządzanie projektami jak nigdy dotąd!"

- **Pierpaolo Vergati**, wykładowca na Uniwersytecie Sapienza w Rzymie i starszy kierownik projektów budowlanych w firmie Fintecna.

"Książkę przeczytałem jednym tchem, w niecałe 6 godzin. Jakość wykonania książki jest doskonała, gęsty błyszczący papier, kolorystyka, przyjemna czcionka. Duża liczba praktycznych przykładów pracy z LLM specyficznych dla branży budowlanej pozwoli zaoszczędzić miesiące, jeśli nie lata, samodzielnej nauki. Przykłady pracy są bardzo zróżnicowane, od prostych do złożonych, bez konieczności zakupu skomplikowanego i drogiego oprogramowania. Książka pozwoli właścicielom każdej firmy z branży budowlanej spojrzeć świeżym okiem na swoją strategię biznesową, cyfryzację i perspektywy rozwoju. A mniejszym firmom zwiększyć wydajność dzięki przystępnom cenowym i darmowym narzędziom".

- **Mikhail Kosarev**, wykładowca i konsultant ds. transformacji cyfrowej w branży budowlanej w TIM-ASG.

Książka "DATA DRIVEN CONSTRUCTION" to przełom dla każdego, kto jest ciekawy, dokąd zmierza branża budowlana w erze danych. Artem nie tylko zarysuje powierzchnię; zagłębia się w bieżące wydarzenia, wyzwania i obiecujące możliwości w budownictwie. To, co wyróżnia tę książkę, to jej przystępność - Artem wyjaśnia złożone idee za pomocą analogii, które sprawiają, że treść jest łatwa do zrozumienia. Uważam, że książka jest niezwykle pouczająca, a jednocześnie wciągająca. Podsumowując, Artem stworzył cenne źródło informacji, które nie tylko informuje, ale także inspiruje. Niezależnie od tego, czy jesteś doświadczonym profesjonalistą, czy nowicjuszem w budownictwie, ta książka poszerzy twoją perspektywę i pogłębi zrozumienie, dokąd zmierza branża. Gorąco polecam!"

- **Moayad Saleh**, architekt i kierownik ds. wdrażania BIM w TMM GROUP Gesamtplanungs GmbH.

"Powiniensem powiedzieć, że Data-Driven Construction jest warta nauczania jako podręcznik na uniwersytetach i jestą książką, która wniesie cenny wkład w rozwijającą się dziedzinę BIM. Książka Data-Driven Construction zawiera słownik techniczny, który bardzo dobrze wyjaśnia pojęcia. Tematy, które są niezwykle trudne do wyjaśnienia, są proste i zrozumiałe dzięki bardzo pięknemu językowi wizualnemu. Uważam, że to, co ma być wyjaśnione w wizualizacjach, powinno być wyrażone czytelnikowi, nawet jeśli w skrócie. Zrozumiałość niektórych wizualizacji, innymi słowy, czytanie wizualizacji wymaga osobnych informacji. Chciałbym również powiedzieć, że z przyjemnością przedstawiam cenne prace Artyoma Boiko podczas moich wykładów i seminariów na uniwersytetach".

- **Dr Ediz Yazicioglu**, właściciel ArchCube i wykładowca zarządzania projektami budowlanymi na Wydziale Architektury Politechniki Stambulskiej i Uniwersytetu Medipol.

"Data Driven Construction" żywo przekazuje podstawy pracy opartej na informacjach z danymi budowlanymi. Książka, która zajmuje się przepływem informacji i podstawowymi koncepcjami ekonomicznymi, a tym samym wyróżnia się na tle innych książek BIM, ponieważ nie tylko reprezentuje perspektywę producenta oprogramowania, ale także stara się przekazać podstawowe koncepcje. Książka warta przeczytania i zobaczenia."

- **Jakob Hirn**, CEO i współzałożyciel Build Informed GmbH oraz inicjator Forum Innowacji "Na szczyt BIM".

"Dane to nowa rupa", jak mówią, więc ich poszukiwacze lub górnicy powinni mieć odpowiednie narzędzia i sposób myślenia, aby wydobyć wartość z tego zasobu XXI wieku. Branża budowlana zbyt długo znajdowała się na ślekim zboczu procesów opartych na "informacjach 3D", w których realizacja projektu opiera się na cudzych upieczętych informacjach (np. już wykreślili wykres kołowy lub słupkowy), podczas gdy podstawowe "dane" (np. surowy arkusz kalkulacyjny) są w stanie dostarczyć znacznie więcej, zwłaszcza że fuzja wielu danych i sztuczna inteligencja niosą ze sobą nieograniczony potencjał. Jeśli zajmujesz się dostarczaniem (lub nauczaniem/badaniami) konstrukcji, ta książka jest najlepszym - i jak dotąd jedynym - źródłem informacji na temat poruszania się po świecie opartym na danych, w którym się znaleźliśmy".

- **Dr Zulfikar Adamu**, profesor nadzwyczajny strategicznego IT w budownictwie na LSBU, Wielka Brytania.

"Data-Driven Construction" autorstwa Artyma Boiko to imponująca praca, która oferuje solidne podstawy dla branży budowlanej w czasach stale rozwijających się technologii i możliwości informacyjnych. Boiko udaje się przedstawić złożone tematy w zrozumiałą sposób, jednocześnie wprowadzając wizjonerskie pomysły. Książka jest dobrze przejętym kompendium, które nie tylko podkreśla obecne osiągnięcia, ale także zapewnia perspektywę przyszłych innowacji. Jest ona wysoce zalecana dla każdego, kto chce poradzić sobie z planowaniem i realizacją budowy w oparciu o dane".

- **Markus Eiberger**, wykładowca na Uniwersytecie Nauk Stosowanych w Stuttgarcie, starszy kierownik projektu i zastępca kierownika oddziału w Konstruktionsgruppe Bauen, członek zarządu stowarzyszenia BIM Cluster Baden-Württemberg.



Drugie wydanie, marzec 2025 r.
© 2025 | Artem Boiko | Karlsruhe

ISBN 978-3-912002-13-3



Artem Boiko Copyright

boikoartem@gmail.com
info@datadrivenconstruction.io

Żadna część tej książki nie może być powielana lub przekazywana w jakiekolwiek formie lub za pomocą jakichkolwiek środków, elektronicznych lub mechanicznych, w tym fotokopii, nagrywania lub jakiegokolwiek systemu przechowywania i wyszukiwania informacji, bez pisemnej zgody autora - z wyjątkiem niekomercyjnej dystrybucji w niezmienionej formie. Książka jest rozpowszechniana bezpłatnie i może być swobodnie udostępniana innym użytkownikom do celów osobistych, edukacyjnych lub badawczych, pod warunkiem zachowania autorstwa i odniesień do oryginału. Autor zachowuje wszelkie niezastrzeżone prawa do tekstu i nie udziela żadnych wyraźnych ani dorozumianych gwarancji. Firmy, produkty i nazwy wymienione w książce mogą być fikcyjne lub użyte jako przykłady. Autor nie ponosi odpowiedzialności za jakiekolwiek konsekwencje wykorzystania podanych informacji. Informacje zawarte w książce są dostarczane "tak jak są" bez gwarancji kompletności lub przydatności. Autor nie ponosi odpowiedzialności za przypadkowe lub wtórne szkody wynikające z wykorzystania informacji, kodu lub programów zawartych w tej książce. Przykłady kodu przedstawione w tej książce służą wyłącznie celom edukacyjnym. Czytelnicy korzystają z nich na własne ryzyko. Autor zaleca przetestowanie wszystkich rozwiązań programowych przed ich użyciem w środowisku produkcyjnym. Wszystkie znaki towarowe i nazwy produktów wymienione w tekście są znakami towarowymi, zastrzeżonymi znakami towarowymi lub znakami usługowymi odpowiednich firm i są własnością ich właścicieli. Użycie tych nazw w książce nie oznacza żadnego związku z ich właścicielami ani ich poparcia. Wymienienie produktów lub usług innych firm nie stanowi rekomendacji ani nie oznacza poparcia. Nazwy firm i produktów użyte w przykładach mogą być znakami towarowymi ich właścicieli. Łącza do stron internetowych osób trzecich zostały zamieszczone dla wygody i nie oznaczają, że autor popiera informacje zawarte na tych stronach. Wszystkie przytoczone statystyki, cytaty i badania były aktualne w momencie ich tworzenia. Dane mogą z czasem ulec zmianie.

Niniejsza książka jest rozpowszechniana na licencji Creative Commons Uznanie autorstwa-Użycie niekomercyjne-Bez utworów zależnych 4.0 Międzynarodowe (CC BY-NC-ND 4.0). Można ją kopiować i rozpowszechniać w celach niekomercyjnych, pod warunkiem zachowania autorstwa i niewprowadzania żadnych zmian.



CC BY-NC-ND 4.0

© 2024 Artyom Boiko. Pierwsze wydanie.
© 2025 Artyom Boiko. Wydanie drugie, poprawione i uzupełnione.
Wszelkie prawa zastrzeżone.

PRZEDMOWA DO DRUGIEGO WYDANIA

Niniejsza książka jest wynikiem ożywionego dialogu ze społecznością zawodową. Opiera się na licznych profesjonalnych dyskusjach na temat zarządzania danymi w branży budowlanej, które miały miejsce na różnych platformach zawodowych i platformach mediów społecznościowych. Dyskusje te stały się podstawą artykułów, publikacji i materiałów wizualnych, które wywołały szeroki odzew w środowisku zawodowym. Treści autora przyciągają miliony wyświetleń każdego roku na różnych platformach i w różnych językach, łącząc profesjonalistów w dziedzinie cyfryzacji budownictwa.

W ciągu roku od publikacji pierwszego wydania książka została zamówiona przez ekspertów z ponad 50 krajów, od Brazylii i Peru po Mauritius i Japonię. Drugie wydanie książki, które trzymasz teraz w rękach, zostało poprawione i rozszerzone w oparciu o opinie ekspertów, krytykę pierwszego wydania i dyskusje w kręgach zawodowych. Dzięki informacjom zwrotnym drugie wydanie zostało znacznie rozszerzone: dodano nowe rozdziały dotyczące technologii CAD (BIM) oraz tworzenia skutecznych procesów ETL. Znaczco zwiększo również liczbę praktycznych przykładów i studiów przypadków. Szczególnie cenne są informacje zwrotne od liderów branży budowlanej, firm konsultingowych i dużych firm informatycznych, którzy zwróciili się do autora z pytaniami dotyczącymi cyfryzacji i interoperacyjności zarówno przed, jak i po opublikowaniu pierwszej wersji książki. Wielu z nich już zastosowało podejścia opisane w książce lub planuje to zrobić w najbliższej przyszłości

Trzymasz w rękach książkę, która powstała w wyniku dyskusji i aktywnej wymiany poglądów. Postępodzi się w dialogu, w ścieraniu się poglądów i otwartości na nowe podejścia. Dziękujemy za udział w tym dialogu. Twoja konstruktywna krytyka jest podstawą przyszłych ulepszeń. Jeśli w tekście występują błędy lub jeśli chciałbyś podzielić się pomysłami i sugestiami, wszelkie informacje zwrotne są mile widziane. Dane kontaktowe znajdują się na końcu książki

DLACZEGO KSIĄŻKA JEST DARMOWA?

Niniejsza książka została pomyślana jako otwarty zasób edukacyjny mający na celu rozpowszechnianie nowoczesnego podejścia do zarządzania danymi w branży budowlanej. Pierwsza wersja książki posłużyła jako podstawa do zebrania komentarzy i sugestii od społeczności zawodowej, co pozwoliło na ulepszenie struktury i treści materiału. Wszystkie komentarze, sugestie i pomysły zostały dokładnie przeanalizowane i włączone do niniejszej poprawionej wersji. Celem książki jest pomoc specjalistom z branży budowlanej w zrozumieniu znaczenia pracy z danymi: systematycznie, świadomie i z myślą o długoterminowej wartości informacji. Autor zebrał przykłady, ilustracje i praktyczne obserwacje z ponad 10 lat pracy nad cyfryzacją budownictwa. Większość tego materiału zrodziła się z rzeczywistych projektów, dyskusji z inżynierami i deweloperami, udziału w międzynarodowych inicjatywach i seminariach szkoleniowych. Książka jest próbą uporządkowania zgromadzonego doświadczenia i podzielenia się nim w przystępny sposób. Jeśli chcesz wesprzeć dalsze rozpowszechnianie pomysłów zawartych w książce i uzyskać wygodny format do czytania, pracy z przykładami i materiałami wizualnymi - możesz kupić [wersję drukowaną](#).

PRAWA UŻYTKOWANIA

Wszystkie materiały, ilustracje i fragmenty niniejszej książki mogą być powielane, cytowane lub wykorzystywane w dowolnym formacie i na dowolnym nośniku, pod warunkiem podania źródła: autorstwa Artyma Boiko i tytułu książki "Data-Driven Construction". Dziękujemy za szacunek dla pracy i rozpowszechnianie wiedzy.

Ze szczerą wdzięcznością dedykuję tę książkę mojej rodzinie, która od najmłodszych lat zaszczepiła we mnie głęboką miłość do budownictwa, mojemu rodzinnemu miastu górnictzemu za lekcje odporności oraz mojej żonie geodecie, której niezawodne wsparcie jest dla mnie nieustającą inspiracją.

DLA KOGO PRZEZNACZONA JEST TA KSIĄŻKA

Napisana przystępnym językiem książka skierowana jest do szerokiego grona czytelników z branży budowlanej - od studentów i nowicjuszy, którzy chcą zrozumieć podstawy nowoczesnych procesów budowlanych, po profesjonalistów, którzy potrzebują aktualnej metodologii zarządzania danymi w budownictwie. Niezależnie od tego, czy jesteś architektem, inżynierem, brygadistą, kierownikiem budowy czy analitykiem danych, ten kompleksowy przewodnik z wieloma unikalnymi ilustracjami i grafikami oferuje cenne informacje na temat wykorzystania danych w biznesie w celu optymalizacji i automatyzacji procesów, usprawnienia procesu podejmowania decyzji i zarządzania projektami budowlanymi na różnych poziomach przy użyciu nowoczesnych narzędzi

Książka jest kompleksowym przewodnikiem, który łączy podstawy teoretyczne i praktyczne zalecenia dotyczące integracji technik zarządzania danymi z procesami budowlanymi. Książka koncentruje się na strategicznym wykorzystaniu informacji w celu optymalizacji operacji, automatyzacji procesów, usprawnienia procesu podejmowania decyzji i skutecznego zarządzania projektami przy użyciu nowoczesnych narzędzi cyfrowych.

Książka ta obejmuje teoretyczne i praktyczne aspekty pracy z informacjami w branży budowlanej. Dzięki szczegółowym przykładom bada metodologię parametryzacji zadań, gromadzenia wymagań, przetwarzania nieustrukturyzowanych i wieloformatowych danych oraz przekształcania ich w skuteczne rozwiązania dla firm budowlanych.

Czytelnik sukcesywnie przechodzi drogę od formułowania wymagań i tworzenia podstawowych modeli danych do bardziej złożonych procesów integracji heterogenicznych źródeł informacji, tworzenia procesów ETL, budowy potoków informacyjnych i modeli uczenia maszynowego. Sekwencyjne podejście pozwala w przejrzysty sposób zademonstrować mechanizmy organizacji i automatyzacji procesów biznesowych oraz systemów wspomagania decyzji w branży budowlanej. Każda część książki kończy się praktycznym rozdziałem zawierającym instrukcje krok po kroku, które umożliwiają natychmiastowe zastosowanie zdobytej wiedzy w rzeczywistych projektach.

STRESZCZENIE CZĘŚCI KSIĄŻKI

Książka ta opiera się na koncepcji transformacji danych w łańcuchu wartości: od gromadzenia danych i zapewniania jakości po przetwarzanie analityczne i wydobywanie cennych praktycznych rozwiązań przy użyciu nowoczesnych narzędzi i metodologii.

Część 1: Cyfrowa ewolucja w budownictwie - śledzi historyczną transformację zarządzania danymi od glinianych tabliczek do nowoczesnych systemów cyfrowych, analizując pojawienie się systemów modułowych i rosnące znaczenie cyfryzacji informacji w kontekście rewolucji przemysłowych.

Część 2: Wyzwania informacyjne dla branży budowlanej - analizuje problemy związane z fragmentacją danych, "silosami informacyjnymi", wpływem podejścia HiPPO na podejmowanie decyzji i ograniczeniami zastrzeżonych formatów, sugerując rozważenie przejścia na ekosystemy AI i LLM.

Część 3: Systematyzacja danych w budownictwie - tworzy typologię danych budowlanych, opisuje metody ich organizacji, integracji z systemami korporacyjnymi oraz omawia tworzenie centrów kompetencyjnych do standaryzacji procesów informacyjnych.

Część 4: Zapewnienie jakości danych - ujawnia metodologie przekształcania rozproszonych informacji w wysokiej jakości, ustrukturyzowane dane, w tym ekstrakcję danych z różnych źródeł, walidację i modelowanie przy użyciu LLM.

Część 5: Kalkulacje kosztów i czasu - dotyczy cyfryzacji kalkulacji kosztów i planowania, automatyzacji uzyskiwania objętości z modeli CAD (BIM), technologii modelowania 4D-8D i kalkulacji ESG projektów budowlanych.

Część 6: CAD i BIM - krytycznie analizuje ewolucję technologii projektowania, kwestie interoperacyjności systemów, trendy w kierunku otwartych formatów danych oraz perspektywy zastosowania sztucznej inteligencji w projektowaniu.

Część 7: Analiza danych i automatyzacja - dotyczy zasad wizualizacji informacji, kluczowych wskaźników wydajności, procesów ETL, narzędzi do orkiestracji przepływu pracy oraz zastosowania modeli językowych do automatyzacji rutynowych zadań.

Część 8: Przechowywanie i zarządzanie danymi - analizuje formaty przechowywania danych, koncepcje hurtowni danych i jezior danych, zasady zarządzania danymi oraz nowe podejścia, w tym wektorowe bazy danych oraz metodologie DataOps i VectorOps.

Część 9: Big Data i uczenie maszynowe - koncentruje się na przejściu do obiektywnej analizy opartej na danych historycznych, Internecie rzeczy na placach budowy oraz zastosowaniu algorytmów uczenia maszynowego do przewidywania kosztów i harmonogramów projektów.

Część 10: Branża budowlana w erze danych cyfrowych - przedstawia spojrzenie na przyszłość branży budowlanej, analizując przejście od analizy przyczynowej do pracy z korelacjami, koncepcję "Uberyzacji" budownictwa i strategie transformacji cyfrowej.

What is meant by **data-driven construction** ?



WPROWADZENIE

Jak długo Twoja firma może pozostać konkurencyjna w świecie, w którym technologia szybko ewoluuje, a każdy aspekt działalności, od harmonogramu i kalkulacji kosztów po analizę ryzyka, jest automatyzowany przez modele uczenia maszynowego?

Branża budowlana, która istnieje tak długo jak sama ludzkość, stoi u progu rewolucyjnych zmian, które zapowiadają całkowitą zmianę sposobu myślenia o tradycyjnym budownictwie. Już teraz w innych sektorach gospodarki cyfryzacja nie tylko zmienia zasady, ale bezwzględnie wypiera z rynku firmy, które nie zdołały dostosować się do nowego środowiska przetwarzania danych i nie są w stanie poprawić szybkości podejmowania decyzji (rys. 1).



Rys. 1 Szybkość podejmowania decyzji w branży budowlanej zależy od czynnika ludzkiego częściej niż w innych branżach.

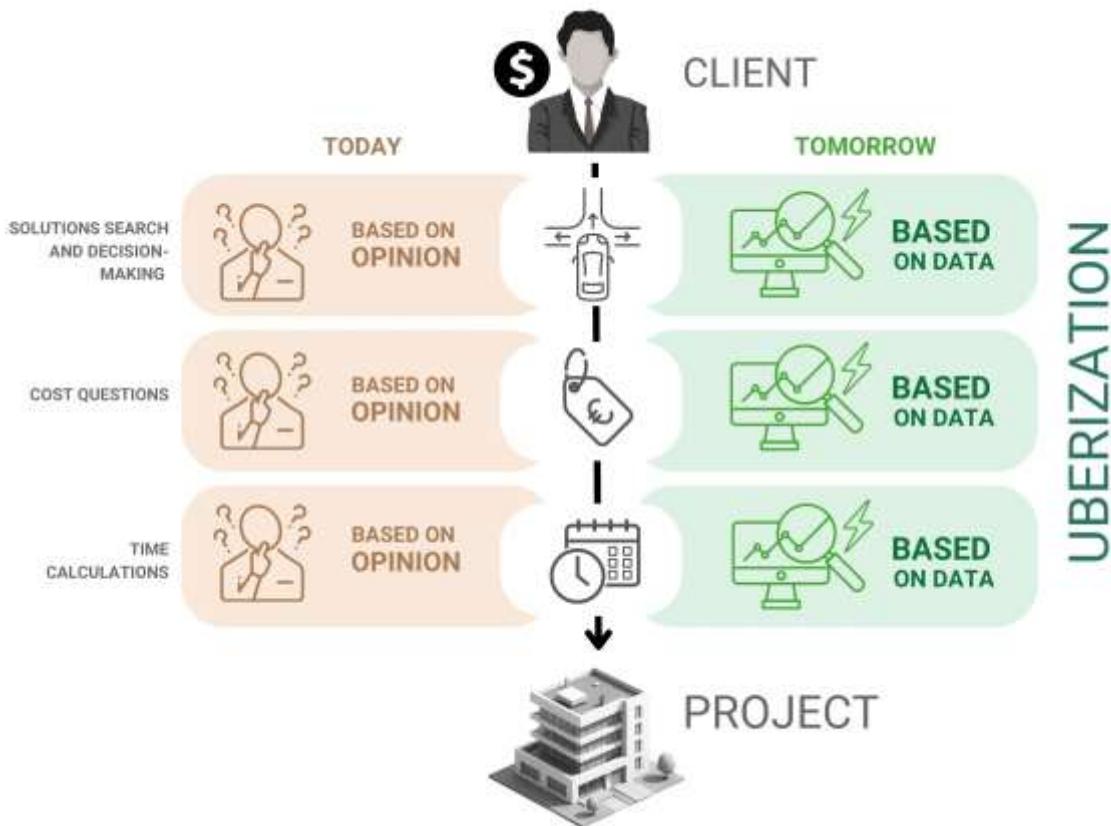
Bankowość, handel detaliczny, logistyka i agrobiznes szybko zmierzają w kierunku pełnej cyfryzacji, w której nie ma już miejsca na nieścisłości i subiektywne opinie. Nowoczesne algorytmy są w stanie analizować ogromne ilości danych i dostarczać klientom dokładnych prognoz - niezależnie od tego, czy chodzi o prawdopodobieństwo spłaty kredytu, optymalne trasy dostaw czy prognozowanie ryzyka.

Branża budowlana jest jedną z ostatnich, które nieuchronnie przechodzą od rozwiązań opartych na opiniach wysoko opłacanych specjalistów do rozwiązań opartych na danych. Przejście to jest napędzane nie tylko przez nowe możliwości technologiczne, ale także przez zwiększone zapotrzebowanie rynku i klientów na przejrzystość, dokładność i szybkość.

Robotyzacja, automatyzacja procesów, otwarte dane i oparte na nich prognozy - wszystko to nie jest już tylko możliwością, ale nieuchronnością. Większość firm z branży budowlanej, które do niedawna były odpowiedzialne przed klientem za obliczanie ilości, kosztów, czasu realizacji projektów i kontroli jakości, teraz ryzykuje przekształcenie się w zwykłych wykonawców zamówień, nie podejmujących kluczowych

decyzji (rys. 2).

Dzięki postępowom w zakresie mocy obliczeniowej, algorytmów uczenia maszynowego i demokratyzacji dostępu do danych, możliwe jest teraz automatyczne łączenie danych z różnych źródeł, umożliwiając głębszą analizę procesów, przewidywanie ryzyka i optymalizację kosztów na etapach dyskusji nad projektem budowlanym. Technologie te mogą radykalnie poprawić wydajność i obniżyć koszty w całym sektorze.



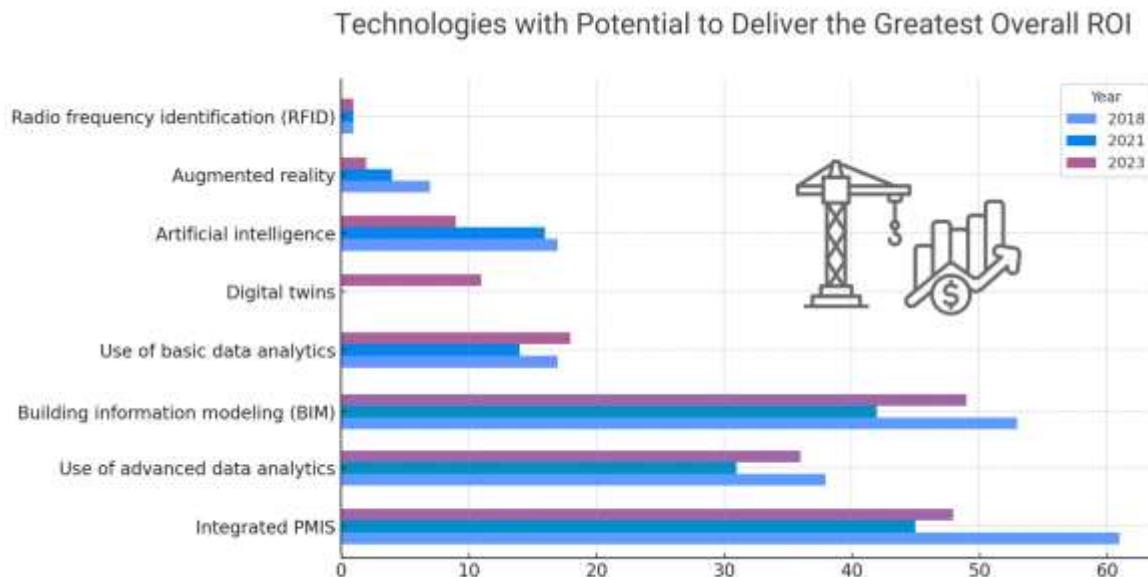
Rys. 2 Klient nie jest zainteresowany nadmierną ilością czynników ludzkich na drodze do realizacji swojego projektu.

Pomimo wszystkich zalet nowych narzędzi i koncepcji, branża budowlana pozostaje daleko w tyle za innymi sektorami gospodarki w zakresie wdrażania nowych technologii.

Według raportu IT Metrics Key Data 2017, branża budowlana zajmuje ostatnie miejsce pod względem wydatków na IT wśród 19 innych branż [1].

Gwałtowny wzrost ilości danych i złożoności procesów staje się bólem głowy dla kierownictwa firmy, a głównym problemem w korzystaniu z nowych technologii jest to, że dane, pomimo ich obfitości, pozostają rozdrobnione, nieuszufluryzowane i często niekompatybilne między różnymi systemami i oprogramowaniem. Dlatego też wiele firm z sektora budowlanego jest obecnie przede wszystkim zaniepokojonych kwestiami jakości danych, które można rozwiązać jedynie poprzez wdrożenie wydajnych, zautomatyzowanych systemów zarządzania i analityki.

Według ankiety przeprowadzonej przez KPMG® wśród kierowników budowy w 2023 roku [2], systemy informatyczne do zarządzania projektami (PMIS), zaawansowana i podstawowa analityka danych oraz modelowanie informacji o budynku (BIM) mają największy potencjał poprawy zwrotu z inwestycji w projekty (Rys. 3).



Rys. 3 Ankieta wśród menedżerów firm budowlanych: które technologie zapewnią najwyższy zwrot z inwestycji (ROI) w projekty kapitałowe? (na podstawie materiałów [2]).

Rozwiązaniem wyzwań związanych z integracją danych z procesami biznesowymi jest zapewnienie wysokiej jakości informacji, wykorzystanie odpowiednich formatów danych i zastosowanie skutecznych metod tworzenia, przechowywania, analizowania i przetwarzania danych.

Uświadomienie sobie wartości danych zmusza różne branże do odejścia od silosowych aplikacji i złożonych biurokratycznych struktur zarządzania. Zamiast tego nacisk przenosi się na tworzenie nowych podejść do architektury informacji, przekształcając firmy w nowoczesne przedsiębiorstwa oparte na danych. Prędzej czy później sama branża budowlana podejmie ten krok, przechodząc od stopniowej ewolucji cyfrowej do prawdziwej rewolucji cyfrowej, która dotknie wszystkie firmy.

Przejście do procesów biznesowych opartych na danych nie będzie łatwe. Wiele firm stanie przed wyzwaniami, ponieważ kadra kierownicza nie zawsze rozumie, jak wykorzystać chaotyczne zbioru danych do poprawy wydajności i rozwoju biznesu.

Ta książka zagłębia się w świat danych, w którym informacje stają się kluczowym zasobem strategicznym, decydującym o wydajności i trwałości procesów biznesowych. Wraz z szybkim wzrostem ilości informacji, firmy stają przed nowymi wyzwaniami. Cyfrowa transformacja nie jest już tylko modnym hasłem - staje się koniecznością.



Rys. 4 Dane i procesy są podstawą konstrukcji.

Zrozumieć transformację oznacza być w stanie wyjaśniać to, co złożone, w prostych słowach. Dlatego też książka napisana jest przystępnym językiem i towarzyszą jej ilustracje autora, stworzone specjalnie w celu jasnego wyjaśnienia kluczowych pojęć. Te diagramy, wykresy i wizualizacje mają na celu usunięcie barier w percepceji i uczynienie materiału zrozumiałym nawet dla tych, którzy wcześniej uważali takie tematy za zbyt złożone. Wszystkie ilustracje, diagramy i grafiki w tej książce zostały stworzone przez autora i zaprojektowane specjalnie w celu wizualizacji kluczowych pojęć opisanych w tekście.

Jeden obraz jest wart tysiąca słów [3].

- Fred R. Barnard, angielski ilustrator, 1927 r.

Aby połączyć teorię z praktyką, wykorzystamy narzędzia sztucznej inteligencji (w szczególności modele językowe), które pozwalają opracowywać rozwiązania bez konieczności posiadania głębokiej wiedzy programistycznej. Jeśli jesteś zorientowany na materiał praktyczny i bardziej interesuje Cię praktyczna praca z danymi, możesz pominąć pierwszą część wprowadzającą i przejść od razu do drugiej części książki, gdzie zaczyna się opis konkretnych przykładów i przypadków.

Nie należy jednak pokładać nadmiernych oczekiwania w narzędziach AI (Artificial Intelligence), uczenia maszynowego i LLM (Large Language Models). Bez wysokiej jakości danych wejściowych i dogłębnego zrozumienia tematu, nawet najbardziej zaawansowane algorytmy nie mogą zapewnić wiarygodnych i znaczących wyników.

Dyrektor generalny Microsoftu Satya Nadella ostrzega przed ryzykiem bańki w sztucznej inteligencji na początku 2025 [4], porównując obecny szum do bańki dot-comów. Podkreśla, że twierdzenia o osiągnięciu kamieni milowych AGI (Artificial General Intelligence) bez odpowiedniego uzasadnienia są "bezsensowną manipulacją wskaźnikami". Nadella uważa, że prawdziwy sukces AI powinien być mierzony jej wkładem we wzrost globalnego PKB, a nie nadmiernym naciskiem na modne hasła.

Za wszystkimi hasłami dotyczącymi nowych technologii i koncepcji kryje się złożona i żmudna praca nad zapewnieniem jakości danych, parametryzacją procesów biznesowych i dostosowaniem narzędzi do rzeczywistych zadań.

Podejście oparte na danych nie jest produktem, który można po prostu pobrać lub kupić. To strategia, którą trzeba zbudować. Zaczyna się od świeżego spojrzenia na istniejące procesy i problemy, a następnie wymaga zdyscyplinowanego ruchu w wybranym kierunku.

Wiodący twórcy oprogramowania i dostawcy aplikacji nie będą motorem zmian w branży budowlanej, ponieważ dla wielu z nich podejście oparte na danych stanowi zagrożenie dla ich ugruntowanego modelu biznesowego.

Inne branże [w przeciwieństwie do budownictwa], takie jak motoryzacja, przeszły już radykalne i przełomowe zmiany, a ich cyfrowa transformacja jest już w toku. Firmy budowlane muszą działać szybko i zdecydowanie: zwinne firmy będą czerpać ogromne korzyści, podczas gdy dla tych, którzy się wahają, ryzyko będzie poważne. Pomyślmy o przewrocie, jaki w tej branży wywołała fotografia cyfrowa [5].

- Raport Światowego Forum Ekonomicznego Shaping the Future of Construction, 2016 r.

Firmy, które w porę dostrzegą możliwości i korzyści płynące z nowego podejścia, zyskają trwałą przewagę konkurencyjną i będą w stanie rozwijać się bez uzależnienia od rozwiązań dużych dostawców.

To Twoja szansa, aby nie tylko przetrwać nadchodzącej burzę cyfryzacji informacji, ale także przejąć nad nią kontrolę. W tej książce znajdziesz nie tylko analizę obecnego stanu branży, ale także konkretne zalecenia dotyczące ponownego przemyślenia i restrukturyzacji procesów i działalności, aby stać się liderem w nowej erze budownictwa i zwiększyć swoje doświadczenie zawodowe.

Cyfrowa przyszłość budownictwa to nie tylko wykorzystanie nowych technologii i programów, ale także fundamentalne przemyślenie obsługi danych i modeli biznesowych.

Czy Twoja firma jest gotowa na tę strategiczną zmianę?

SPIS TREŚCI

WPROWADZENIE.....	1
SPIS TREŚCI	I
II CZĘŚĆ OD GLINIANYCH TABLICZEK DO CYFROWEJ REWOLUCJI: JAK EWOLUOWAŁA INFORMACJA W BUDOWNICTWIE	2
ROZDZIAŁ 2.1. EWOLUCJA WYKORZYSTANIA DANYCH W BRANŻY BUDOWLANEJ	3
Narodziny ery danych w budownictwie.....	3
Od gliny i papirusu do technologii cyfrowej.....	4
Proces jako narzędzie do tworzenia doświadczeń opartych na danych	5
Cyfryzacja informacji o procesach budowlanych.....	7
ROZDZIAŁ 2.2. TECHNOLOGIE I SYSTEMY ZARZĄDZANIA W NOWOCZESNYM BUDOWNICTWIE	11
Rewolucja cyfrowa i pojawienie się modułowych systemów MRP/ERP -systems.....	11
Systemy zarządzania danymi: od eksploracji danych do wyzwań biznesowych.....	13
Korporacyjna grzybnia: jak dane łączą się z procesami biznesowymi	16
ROZDZIAŁ 2.3. REWOLUCJA CYFROWA I EKSPLOZJA DANYCH	20
Początek boomu na dane jako fala ewolucyjna	20
Ilość danych generowanych w nowoczesnej firmie.....	22
Koszt przechowywania danych: aspekt ekonomiczny	23
Granice gromadzenia danych: od masy do znaczenia	25
Kolejne kroki: od teorii danych do praktycznych zmian	27
ROZDZIAŁ 2.4. JAK BRANŻA BUDOWLANA TONIE W CHAOSIE DANYCH.....	29
ROZDZIAŁ 2.5. FRAGMENTACJA I SILOSY DANYCH.....	30
Im więcej narzędzi, tym bardziej wydajny biznes?	30
Silosy danych i ich wpływ na wydajność firmy.....	32
Powielanie i brak jakości danych jako konsekwencja braku jedności.....	35
HiPPO lub niebezpieczeństwo opinii w podejmowaniu decyzji.....	37
Ciągły wzrost złożoności i dynamiki procesów biznesowych	40
Czwarta rewolucja przemysłowa (Przemysł 4.0) i piąta rewolucja przemysłowa (Przemysł 5.0) w budownictwie.....	43
ROZDZIAŁ 2.6. PRZEKSZTAŁCANIE CHAOSU W PORZĄDEK I ZMNIEJSZANIE ZŁOŻONOŚCI.....	46
Zbędny kod i zamknięte systemy jako bariera dla poprawy produktywności	46
Od silosów do pojedynczej hurtowni danych	48

Zintegrowane systemy pamięci masowej umożliwiają przejście na agentów AI.....	49
Od gromadzenia danych do podejmowania decyzji: droga do automatyzacji	52
Kolejne kroki: przekształcenie chaosu w zarządzalny system.....	54
III CZĘŚĆ STRUKTURA DANYCH W PROCESACH BIZNESOWYCH W BRANŻY BUDOWLANEJ 56	
ROZDZIAŁ 3.1. TYPY DANYCH W BUDOWNICTWIE	57
Najważniejsze typy danych w branży budowlanej.....	57
Dane strukturalne	61
Relacyjne bazy danych RDBMS i język zapytań SQL.....	62
SQL - zapytania w bazach danych i nowe trendy	65
Dane niestrukturyzowane	67
Dane tekstowe: między chaosem niestrukturyzowanym a chaosem ustrukturyzowanym y	68
Dane częściowo ustrukturyzowane i luźno ustrukturyzowane.....	70
Dane geometryczne i ich zastosowanie	71
Dane CAD: od projektu do przechowywania danych.....	73
Pojawienie się koncepcji BIM (BOM) i wykorzystanie CAD w procesach	76
ROZDZIAŁ 3.2. HARMONIZACJA I STRUKTURYZACJA DANYCH..... 82	
Wypełnianie systemów danymi w branży budowlanej.....	82
Transformacja danych: kluczowy fundament nowoczesnej analizy biznesowej	85
Modele danych: relacje w danych i relacje między elementami	88
Zastrzeżone formaty i ich wpływ na procesy cyfrowe	92
Otwarte formaty zmieniają podejście do cyfryzacji.....	96
Zmiana paradygmatu: Open Source jako koniec ery dominacji producentów oprogramowania	97
Ustrukturyzowane otwarte dane: fundament cyfrowej transformacji.....	100
ROZDZIAŁ 3.3. LLM I JEGO ROLA W PRZETWARZANIU DANYCH I PROCESACH BIZNESOWYCH	103
LLM czytają: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok do automatyzacji procesów danych	103
Duże modele językowe LLM: jak to działa?.....	104
Korzystanie z lokalnych usług LLM dla wrażliwych danych firmowych.....	107
Pełna kontrola nad AI w firmie i jak wdrożyć własne LLM	109
RAG: Inteligentny LLM - asystenci z dostępem do danych korporacyjnych.....	111
ROZDZIAŁ 3.4. IDE Z OBSŁUGĄ LLM I PRZYSZŁYMI ZMIANAMI W PROGRAMOWANIU..... 113	
Wybór IDE: od eksperymentów LLM do rozwiązań biznesowych.....	113
IDE z obsługą LLM i przyszłe zmiany w programowaniu	115
Python Pandas: niezastąpione narzędzie do pracy z danymi.....	116

DataFrame: uniwersalny format danych tabelarycznych	120
Kolejne kroki: budowanie zrównoważonych ram danych	123
IV CZĘŚĆ JAKOŚĆ DANYCH: ORGANIZACJA, STRUKTURA, MODELOWANIE 125	
ROZDZIAŁ 4.1. KONWERSJA DANYCH DO POSTACI USTRUKTURYZOWANEJ 126	
Nauka przekształcania dokumentów, plików PDF, obrazów i tekstów w ustrukturyzowane formaty	126
Przykład konwersji dokumentu PDF na tabelę	127
Konwersja obrazów JPEG, PNG do postaci strukturalnej.....	131
Konwersja danych tekstowych do postaci ustrukturyzowanej.....	134
Konwersja danych CAD (BIM) do postaci ustrukturyzowanej	137
Dostawcy rozwiązań CAD zmierzają w kierunku ustrukturyzowanych danych	142
ROZDZIAŁ 4.2. KLASYFIKACJA I INTEGRACJA: WSPÓLNY JĘZYK DLA DANYCH BUDOWLANYCH	145
Szybkość podejmowania decyzji zależy od jakości danych.....	145
Standaryzacja i integracja	146
Interoperacyjność cyfrowa zaczyna się od wymagań	149
Wspólny język konstrukcji: rola klasyfikatorów w transformacji cyfrowej.....	151
Masterformat, OmniClass, Uniclass i CoClass: ewolucja systemów klasyfikacji.....	154
ROZDZIAŁ 4.3. MODELOWANIE DANYCH I CENTRUM DOSKONAŁOŚCI	159
Modelowanie danych: model koncepcyjny, logiczny i fizyczny	159
Praktyczne modelowanie danych w kontekście budownictwa	162
Tworzenie bazy danych przy użyciu LLM	165
Centrum doskonałości (CoE) w zakresie modelowania danych.....	167
ROZDZIAŁ 4.4. USYSTEMATYZOWANIE WYMAGAŃ I WALIDACJA INFORMACJI	170
Gromadzenie i analiza wymagań: przekształca komunikację w ustrukturyzowane dane	170
Schematy blokowe procesów i skuteczność ram koncepcyjnych.....	174
Wymagania strukturalne i wyrażenia regularne RegEx	176
Gromadzenie danych dla procesu weryfikacji	181
Weryfikacja danych i wyniki weryfikacji	183
Wizualizacja wyników weryfikacji	188
Porównanie kontroli jakości danych z potrzebami ludzkiego życia	190
Kolejne kroki: przekształcanie danych w dokładne obliczenia i plany.....	192
V CZĘŚĆ KALKULACJE KOSZTÓW I CZASU: WŁĄCZANIE DANYCH DO PROCESÓW BUDOWLANYCH 194	
ROZDZIAŁ 5.1. KALKULACJE KOSZTÓW I SZACUNKI DLA PROJEKTÓW BUDOWLANYCH	195

Podstawy budownictwa: szacowanie ilości, kosztów i czasu.....	195
Metody obliczania szacunkowych kosztów projektów	196
Metoda oparta na zasobach kosztorysowanie i szacowanie w budownictwie	197
Baza danych zasobów budowlanych: katalog materiałów i robót budowlanych	197
Kompilacja obliczeń i kosztorysowanie prac w oparciu o bazę zasobów	199
Ostateczny kosztorys projektu: od szacunków do budżetów	204
ROZDZIAŁ 5.2. PRZEDMIAR I AUTOMATYCZNE TWORZENIE KOSZTORYSÓW I HARMONOGRAMÓW	208
Przejście z 3D do 4D i 5D: wykorzystanie parametrów objętościowych i ilościowych	208
Atrybuty 5D i pobieranie objętości atrybutów z CAD	208
QTO Quantity Take-Off: grupowanie danych projektu według atrybutów	212
Automatyzacja QTO przy użyciu LLM i danych strukturalnych	217
QTO obliczanie całego projektu przy użyciu reguł grupowych z arkusza kalkulacyjnego Excel.....	221
ROZDZIAŁ 5.3. 4D, 6D -8D I OBLCZENIA EMISJI DWUTLENKU WĘGLA CO₂.....	227
Model 4D: uwzględnienie czasu w szacunkach konstrukcyjnych	227
Harmonogram budowy i jego automatyzacja w oparciu o dane kosztorysowe	228
Rozszerzone warstwy atrybutów 6D -8D: od efektywności energetycznej do zapewnienia bezpieczeństwa	230
Szacowanie emisji CO ₂ i obliczanie emisji dwutlenku węgla z projektów budowlanych	233
ROZDZIAŁ 5.4. SYSTEMY BUDOWLANE ERP I PMIS	237
Construction ERP -systemy na przykładzie kalkulacji i kosztorysów.....	237
PMIS: Pośrednik między ERP a placem budowy	242
Spekulacja, zysk, wyspiarski charakter i brak przejrzystości w ERP i PMIS	243
Koniec ery zamkniętych systemów ERP /PMIS: branża budowlana potrzebuje nowego podejścia..	245
Kolejne kroki: efektywne wykorzystanie danych projektowych.....	247
VI CZĘŚĆ CAD I BIM: MARKETING, RZECZYWISTOŚĆ I PRZYSZŁOŚĆ DANYCH PROJEKTOWYCH W BUDOWNICTWIE	250
ROZDZIAŁ 6.1. POJAWIENIE SIĘ KONCEPCJI BIM- W BRANŻY BUDOWLANEJ.....	251
Historia pojawienia się BIM i otwartego BIM jako koncepcji marketingowych dostawców CAD-	251
Rzeczywistość BIM: zamiast zintegrowanych baz danych - zamknięte systemy modułowe	254
Pojawienie się otwartego formatu IFC w branży budowlanej	256
Problem z formatem IFC w zależności od rdzenia geometrycznego.....	258
Pojawienie się w konstrukcji tematu semantyki i ontologii.....	261
Dlaczego technologie semantyczne nie spełniają oczekiwania branży budowlanej?	263
ROZDZIAŁ 6.2. ZAMKNIĘTE FORMATY PROJEKTÓW I KWESTIE INTEROPERACYJNOŚCI.....	267

Zamknięte dane i spadająca produktywność: ślepy zaułek branży CAD (BIM).....	267
Mit interoperacyjności między systemami CAD	269
Przejdź do USD i danych granularnych m	273
ROZDZIAŁ 6.3. GEOMETRIA W BUDOWNICTWIE: OD LINII DO METRÓW SZEŚCIENNYCH	277
Kiedy linie zamieniają się w pieniądze lub dlaczego budowniczowie potrzebują geometrii.....	277
Od linii do objętości: Jak powierzchnia i objętość stają się danymi.....	277
Przejście na MESH, USD i wielokąty: korzystanie z teselacji dla geometrii	280
LOD, LOI, LOMD - unikalna klasyfikacja detali w CAD (BIM)	281
Nowe standardy CAD (BIM) - AIA, BEP, IDS, LOD, COBie	284
ROZDZIAŁ 6.4. PARAMETRYZACJA PROJEKTU I WYKORZYSTANIE LLM DO OBSŁUGI CAD	288
Iluzja unikalności danych CAD (BIM): droga do analityki i otwartych formatów.....	288
Projektowanie poprzez parametry: przyszłość CAD i BIM	291
Pojawienie się LLM w projektowaniu CAD procesy przetwarzania danych.....	294
Zautomatyzowana analiza plików DWG za pomocą LLM i Pandas	297
Kolejne kroki: przejście od zamkniętych formatów do otwartych danych.....	303
VII CZĘŚĆ PODEJMOWANIE DECYZJI W OPARCIU O DANE, ANALITYKA, AUTOMATYZACJA I UCZENIE MASZYNOWE	305
ROZDZIAŁ 7.1. ANALITYKA DANYCH I PODEJMOWANIE DECYZJI W OPARCIU O DANE	306
Dane jako zasób w procesie podejmowania decyzji.....	306
Wizualizacja danych: klucz do zrozumienia i podejmowania decyzji	309
KPI i ROI.....	311
Dashboardy i pulpity nawigacyjne: wizualizacja wskaźników w celu skutecznego zarządzania.....	313
Analiza danych i sztuka zadawania pytań	315
ROZDZIAŁ 7.2. PRZEPŁYW DANYCH BEZ RĘCZNEGO WYSIŁKU: DLACZEGO ETL JEST POTRZEBNY	318
Automatyzacja ETL: redukcja kosztów i przyspieszenie obsługi danych	318
ETL Extract: gromadzenie danych	322
ETL Transform: zastosowanie reguł walidacji i transformacji.....	324
ETL Load: Wizualizacja wyników na wykresach i grafach	327
ETL Load: Automatyczne tworzenie dokumentów PDF	332
ETL Load: automatyczne generowanie dokumentów z FPDF	333
ETL Load: Raportowanie i ładowanie do innych systemów	337
ETL z LLM: Wizualizacja danych z dokumentów PDF	338
ROZDZIAŁ 7.3. AUTOMATYCZNY PRZENOŚNIK ETL (RUROCIAĞ)	343
Pipeline: Automatyczny przenośnik ETL danych	343

Pipeline -ETL proces walidacji danych z LLM.....	347
Rurociąg -ETL: weryfikacja danych i informacji o elementach projektu w CAD (BIM).....	349
ROZDZIAŁ 7.4. ORKIESTRACJA ETL I PRZEPŁYWÓW PRACY: PRAKTYCZNE ROZWIĄZANIA.....	355
DAG i Apache Airflow: automatyzacja i orkiestracja przepływu pracy	355
Apache Airflow: praktyczne zastosowanie w automatyzacji ETL.....	356
Apache NiFi do routingu i konwersji danych.....	360
n8n Low-Code, No-Code orkiestracja procesów.....	361
Kolejne kroki: przejście od operacji ręcznych do rozwiązań opartych na analityce	364
VIII CZĘŚĆ PRZECHOWYWANIE I ZARZĄDZANIE DANYMI W BUDOWNICTWIE	366
ROZDZIAŁ 8.1. INFRASTRUKTURA DANYCH: OD FORMATÓW PRZECHOWYWANIA DO REPOZYTORIÓW	367
Atomy danych: podstawa skutecznego zarządzania informacjami	367
Przechowywanie informacji: pliki lub dane	368
Przechowywanie dużych ilości danych: analiza popularnych formatów i ich skuteczności	370
Optymalizacja pamięci masowej za pomocą Apache Parquet.....	373
DWH: Hurtownia danych hurtownie danych	375
Data Lake - ewolucja ETL do ELT: od tradycyjnego czyszczenia do elastycznego przetwarzania	377
Architektura Data Lakehouse: synergia hurtowni i jezior danych	379
CDE, PMIS, ERP lub DWH i Data Lake	381
ROZDZIAŁ 8.2. ZARZĄDZANIE HURTOWNIĄ DANYCH I ZAPOBIEGANIE CHAOSOWI	385
Wektorowe bazy danych i pole ograniczające	385
Data Governance), Data Minimalism) i Data Swamp)	388
DataOps i VectorOps: nowe standardy danych	391
Kolejne kroki: od chaotycznej pamięci masowej do uporządkowanej pamięci masowej	393
IX CZĘŚĆ DUŻE ZBIORY DANYCH, UCZENIE MASZYNOWE I PROGNOZY	395
ROZDZIAŁ 9.1. BIG DATA I ICH ANALIZA	396
Big Data w budownictwie: od intuicji do przewidywalności.....	396
Kwestionowanie wykonalności dużych zbiorów danych: korelacja, statystyka i próbkowanie danych.....	397
Big data: analiza danych ze zbioru milionów pozwoleń na budowę w San Francisco	400
Przykład big data w oparciu o dane CAD (BIM).....	406
Internet rzeczy i inteligentne kontrakty IoT.....	410
ROZDZIAŁ 9.2. UCZENIE MASZYNOWE I PROGNOZY	414
Uczenie maszynowe i sztuczna inteligencja zmienią sposób, w jaki budujemy	414

Od subiektywnej oceny do statystycznej prognozy	416
Zbiór danych Titanic: Witaj świecie w świecie analityki dane i duże zbiory danych.....	418
Uczenie maszynowe w akcji: od pasażerów Titanica do zarządzania projektami	423
Przewidywania i prognozy oparte na danych historycznych	427
Kluczowe koncepcje uczenia maszynowego	429
ROZDZIAŁ 9.3. PRZEWIDYWANIE KOSZTÓW I TERMINÓW PRZY UŻYCIU UCZENIA MASZYNOWEGO	432
Przykład wykorzystania uczenia maszynowego do znalezienia kosztów i harmonogramów projektu.....	432
Przewidywanie kosztów i czasu projektu przy użyciu regresji liniowej	434
Prognozy kosztów i czasu projektu przy użyciu algorytmu K-najbliższych sąsiadów (k-NN)	437
Kolejne kroki: od przechowywania do analizy i prognozowania	441
X CZĘŚĆ BRANŻA BUDOWLANA W ERZE CYFROWEJ. MOŻLIWOŚCI I WYZWANIA.....	444
ROZDZIAŁ 10.1. STRATEGIE PRZETRWANIA: BUDOWANIE PRZEWAGI KONKURENCYJNEJ.....	445
Korelacje zamiast obliczeń: przyszłość analityki budowlanej	445
Podejście oparte na danych w budownictwie: nowy poziom infrastruktury	448
Cyfrowe biuro nowej generacji: jak sztuczna inteligencja zmienia przestrzeń roboczą	450
Otwarte dane i Uberyzacja stanowią zagrożenie dla istniejącej działalności budowlanej	452
Nierozwiążane problemy uberyzacji ostatnią szansą na wykorzystanie czasu na transformację ..	455
ROZDZIAŁ 10.2. PRAKTYCZNY PRZEWODNIK PO WDRAŻANIU PODEJŚCIA OPARTEGO NA DANYCH..	460
Od teorii do praktyki: mapa drogowa cyfrowej transformacji w budownictwie	460
Tworzenie cyfrowych fundamentów: 1-5 kroków do cyfrowej dojrzałości.....	462
Uwolnienie potencjału danych: 5-10 kroków do cyfrowej dojrzałości	466
Plan transformacji: od chaosu do firmy opartej na danych.....	473
Budowanie w Przemyśle 5.0: Jak zarabiać pieniądze, gdy nie można się już ukryć?.....	476
WNIOSKI	478
O AUTORZE	481
DALSZE RELACJE	482
KOMENTARZ DO TŁUMACZENIA	482
INNE UMIEJĘTNOŚCI I POJĘCIA.....	483
SŁOWNICZEK	487
LISTA REFERENCJI I MATERIAŁÓW ONLINE	493
INDEKS PRZEDMIOTÓW	509

MAKSYMALNA WYGODA DZIĘKI WERSJI DRUKOWANEJ

Posiadasz darmową wersję cyfrową **Data-Driven Construction**. W celu wygodniejszej pracy i szybkiego dostępu do materiałów zalecamy zwrócenie uwagi na **wydanie drukowane**:



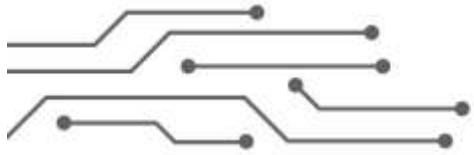
■ **Zawsze pod ręką:** książka w formacie drukowanym stanie się niezawodnym narzędziem pracy, pozwalającym szybko znaleźć i wykorzystać odpowiednie wizualizacje i diagramy w każdej sytuacji roboczej.

■ **Wysokiej jakości ilustracje:** wszystkie obrazy i grafiki w drukowanym wydaniu są prezentowane w maksymalnej jakości.

■ **Szybki dostęp do informacji:** łatwa nawigacja, możliwość tworzenia notatek, zakładek i pracy z książką w dowolnym miejscu.

Kupując pełną drukowaną wersję książki, otrzymujesz wygodne narzędzie do wygodnej i efektywnej pracy z informacją: możliwość szybkiego korzystania z materiałów wizualnych w codziennych zadaniach, szybkiego wyszukiwania potrzebnych schematów i robienia notatek. Ponadto zakup wspiera rozpowszechnianie otwartej wiedzy.

Zamów drukowaną wersję książki na stronie: datadrivenconstruction.io/books



I CZĘŚĆ

OD GLINIANYCH TABLICZEK DO CYFROWEJ REWOLUCJI: JAK EWOLUOWAŁA INFORMACJA W BUDOWNICTWIE

Pierwsza część książki analizuje historyczną ewolucję zarządzania danymi w branży budowlanej, od prymitywnych zapisów na nośnikach fizycznych po nowoczesne ekosystemy cyfrowe. Analizuje transformację technologii zarządzania informacją, pojawienie się systemów ERP oraz wpływ fragmentacji danych na efektywność procesów biznesowych. Szczególną uwagę poświęcono procesowi cyfryzacji informacji i rosnącemu znaczeniu obiektywnej analizy w miejsce subiektywnej oceny ekspertów. Szczegółowo przeanalizowano wykładowczy wzrost ilości informacji, z jakim ma do czynienia współczesna branża budowlana i związane z tym wyzwania dla systemów korporacyjnych. Analizowane jest również pozycjonowanie branży budowlanej w kontekście czwartej i piątej rewolucji przemysłowej, a także potencjał wykorzystania sztucznej inteligencji i podejścia skoncentrowanego na danych w celu stworzenia trwałej przewagi konkurencyjnej.

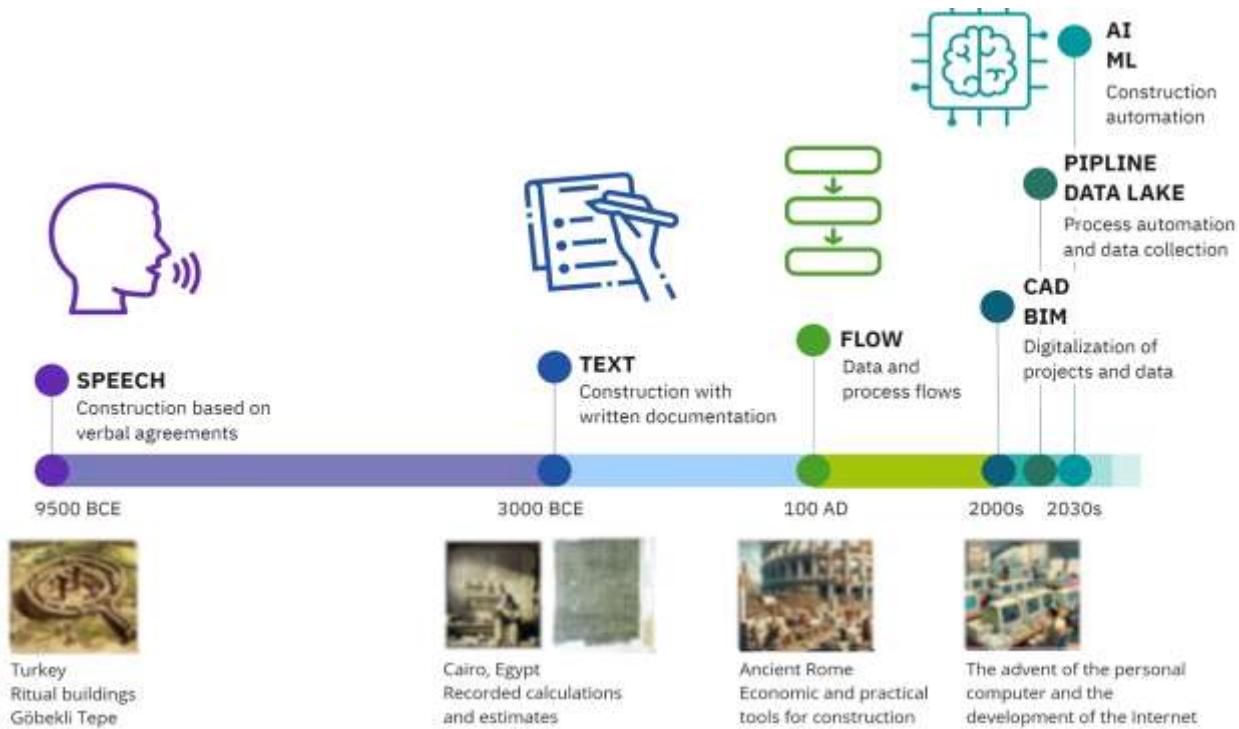
ROZDZIAŁ 1.1.

EWOLUCJA WYKORZYSTANIA DANYCH W BRANŻY BUDOWLANEJ

Narodziny ery danych w budownictwie

Około 10 000 lat temu, w okresie neolitu, ludzkość dokonała rewolucyjnego przejścia w swoim rozwoju, porzucając koczowniczy tryb życia na rzecz osiadłego, co doprowadziło do pojawienia się pierwszych prymitywnych budynków wykonanych z gliny, drewna i kamienia [6]. Od tego momentu rozpoczyna się historia przemysłu budowlanego.

Wraz z rozwojem cywilizacji architektura stawała się coraz bardziej złożona, co doprowadziło do powstania pierwszych rytualnych świątyń i budynków użyteczności publicznej. Rosnąca złożoność projektów architektonicznych wymagała od inżynierów i menedżerów starożytności stworzenia pierwszych zapisów i obliczeń. Pierwsze zapisy na glinianych tabliczkach i papirusach często zawierały opis logiki obliczania ilości potrzebnych materiałów budowlanych, ich kosztów i obliczania płatności za wykonaną pracę [7]. W ten sposób rozpoczęła się era wykorzystywania danych w budownictwie - na dłużej przed pojawieniem się nowoczesnych technologii cyfrowych (Rys. 1.1-1).

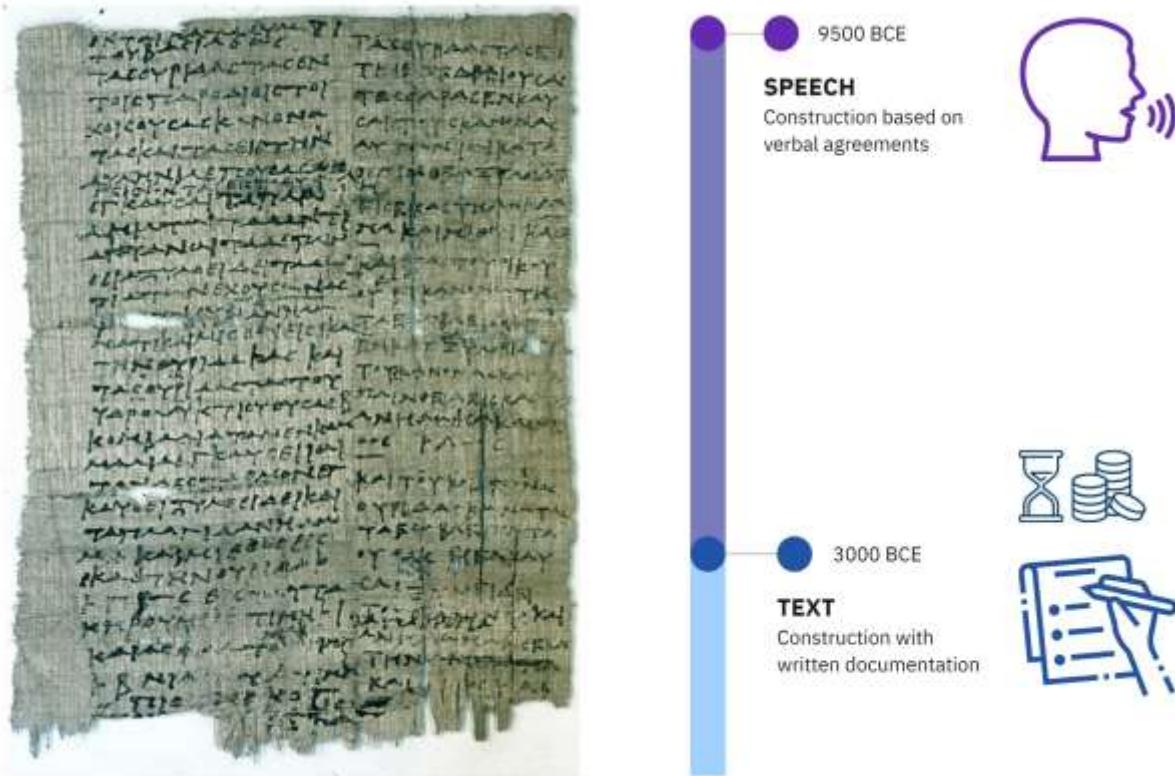


Rys. 1.1-1 Chronologia rozwoju technologii informatycznych w budownictwie: od informacji werbalnej do sztucznej inteligencji.

Od gliny i papirusu do technologii cyfrowej

Pierwsze dokumenty dotyczące budownictwa pochodzą z okresu budowy piramid, około 3000-4000 lat p.n.e.[7]. Od tego czasu prowadzenie pisemnych zapisów ułatwiło i towarzyszyło postępowi w branży budowlanej, umożliwiając gromadzenie i systematyzację wiedzy, która w ciągu kolejnych 10 000 lat doprowadziła do znaczących innowacji w metodach budowlanych i architekturze.

Wykorzystanie pierwszych fizycznych nośników w budownictwie, takich jak gliniane tabliczki, papirus sprzed tysiąc lat (Rys. 1.1-2) lub papier "A0" w latach 80-tych XX wieku, do rejestrowania danych nie miało pierwotnie na celu zastosowania tych informacji w nowych projektach. Głównym celem takich zapisów było wyszczególnienie aktualnego stanu projektu, w tym obliczenia wymaganych materiałów i kosztów prac. Podobnie w dzisiejszym świecie, dostępność cyfrowych danych projektowych i modeli nie zawsze gwarantuje ich zastosowanie w przyszłych projektach i często służy głównie jako informacja do bieżących obliczeń wymaganych materiałów i kosztów budowy.



Rys. 1.1-2 Papirus z III w. p.n.e. opisujący koszty malowania różnych typów okien w pałacu królewskim techniką enkaustyczną.

Ludzkość potrzebowała około 5000 lat, aby przejść od ustnych rozmów do pisemnych dokumentów w zarządzaniu projektami budowlanymi i tyle samo czasu, aby przejść od papieru do danych cyfrowych jako podstawowego zasobu do planowania i kontroli.

Podobnie jak rozwój handlu i stosunków pieniężnych stymulował pojawienie się pisma i pierwszych prawników do rozstrzygania sporów, tak pierwsze zapisy kosztów materiałów i zakresów prac w budownictwie doprowadziły do pojawienia się pierwszych menedżerów w branży budowlanej, których obowiązki obejmowały dokumentowanie, monitorowanie i odpowiedzialność za kluczowe informacje o harmonogramach i kosztach projektów.

Obecnie dane odgrywają znacznie ważniejszą rolę: nie tylko rejestrują podjęte decyzje, ale także stają się narzędziem do przewidywania i modelowania przyszłości. Jest to fundament, na którym opiera się nowoczesne podejście procesowe w zarządzaniu projektami - przekształcenie zgromadzonego doświadczenia w system podejmowania decyzji oparty na ustrukturyzowanych i weryfikowalnych danych.

Proces jako narzędzie do tworzenia doświadczeń opartych na danych

Sercem każdego procesu jest przekształcenie przeszłych doświadczeń w narzędzie do planowania przyszłości. Doświadczenie w nowoczesnym sensie to ustrukturyzowany zestaw danych, które można analizować w celu dokonywania świadomych prognoz.

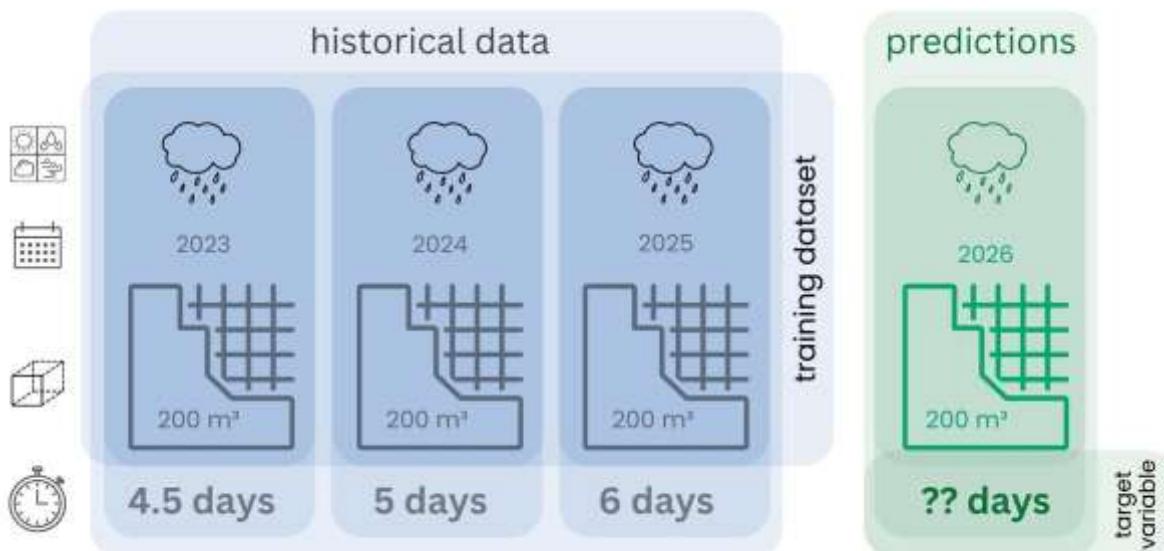
To właśnie dane historyczne służą jako podstawa prognozowania, ponieważ wyraźnie pokazują wyniki wykonanej pracy i zapewniają wgląd w czynniki wpływające na te wyniki.

Weźmy konkretny przykład z budownictwa monolitycznego: zwykle przy planowaniu harmonogramu prac bierze się pod uwagę objętość betonu, złożoność konstrukcji i warunki pogodowe. Założymy, że konkretny brygadzista budowy lub dane historyczne firmy z ostatnich trzech lat (2023-2025) pokazują, że wyłanie monolitycznej konstrukcji o powierzchni 200 m² w deszczową pogodę zajęło od 4,5 do 6 dni (rys. 1.1-3). To właśnie te zgromadzone statystyki stają się podstawą do przewidywania czasu realizacji i kosztów zasobów podczas planowania podobnych prac w przyszłych projektach. Na podstawie tych danych historycznych brygadzista lub kosztorysant może dokonać świadomiej prognozy, opartej na doświadczeniu, czasu wymaganego do ukończenia podobnych prac w przyszłości w 2026 r. w podobnych warunkach.

W przypadku analityki czasu proces analityczny działa jako mechanizm przekształcania rozproszonych danych w ustrukturyzowane doświadczenia, a następnie w precyzyjne narzędzie planowania. Dane i procesy stanowią jeden ekosystem, w którym jedno nie może istnieć bez drugiego.

Policz to, co da się policzyć, zmierz to, co da się zmierzyć, a to, czego nie da się zmierzyć, uczyń mierzalnym [8].

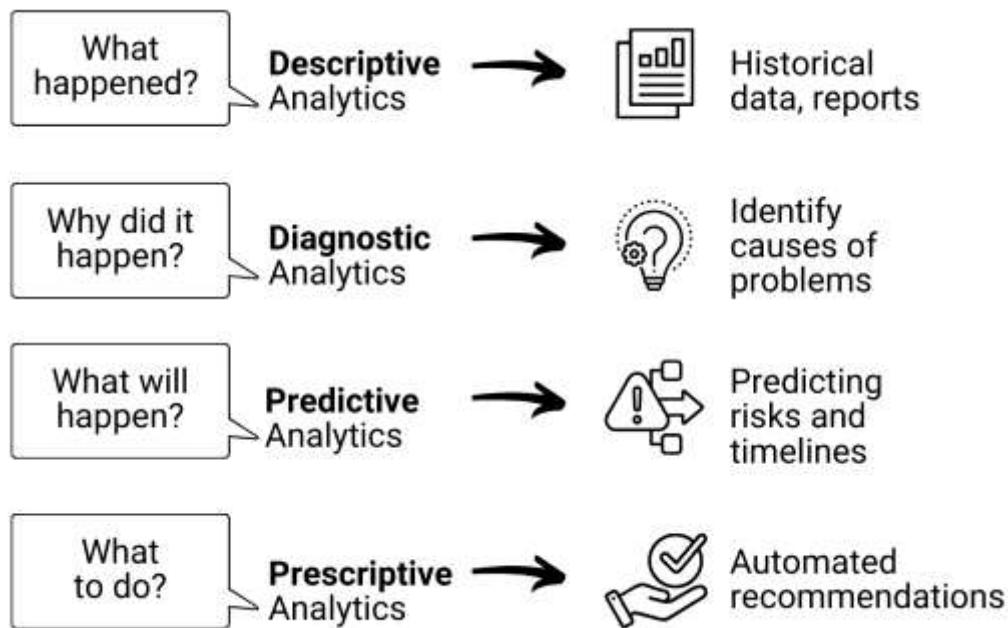
- Galileo Galilei



Rys. 1.1-3 Dane historyczne działają jako zbiór danych szkoleniowych do przewidywania jednej z wartości w przyszłości.

W dzisiejszym krajobrazie biznesowym analityka danych staje się kluczowym elementem skutecznego zarządzania projektami, optymalizacji procesów i podejmowania strategicznych decyzji. Branża budowlana stopniowo opanowuje cztery kluczowe poziomy analityki, z których każdy odpowiada na konkretne pytanie i zapewnia wyjątkowe korzyści (rys. 1.1-4):

- **Analityka opisowa** - odpowiada na pytanie "co się stało?" i dostarcza dane historyczne oraz raporty dotyczące przeszłych zdarzeń i wyników: w ciągu ostatnich trzech lat (2023-2025) wywanie monolitycznej konstrukcji o powierzchni 200 m² przy deszczowej pogodzie zajęło od 4,5 do 6 dni.
- **Analityka diagnostyczna** - odpowiada na pytanie "dlaczego tak się stało?", identyfikując przyczyny problemów: analiza pokazuje, że czas wylewania konstrukcji monolitycznej wydłużył się z powodu deszczowej pogody, która spowolniła proces utwardzania betonu.
- **Analityka predykcyjna** - wybiegająca w przyszłość, przewidująca możliwe ryzyko i czas realizacji poprzez odpowiedź na pytanie "co się stanie?": na podstawie danych historycznych przewiduje się, że wywanie podobnej monolitycznej konstrukcji o powierzchni 200 m² w deszczową pogodę w 2026 r. zajmie około 5,5 dnia, biorąc pod uwagę wszystkie znane czynniki i trendy.
- **Analityka preskryptywna** - zapewnia zautomatyzowane rekomendacje i odpowiada na pytanie "co robić?", umożliwiając firmom wybór optymalnych działań: Aby zoptymalizować prace, zaleca się na przykład: stosowanie specjalnych dodatków przyspieszających utwardzanie betonu w warunkach wysokiej wilgotności; planowanie wylewania na okresy o najmniejszym prawdopodobieństwie wystąpienia opadów; zorganizowanie tymczasowych schronień dla konstrukcji, co pozwoli skrócić czas prac do 4-4,5 dnia nawet w niesprzyjających warunkach pogodowych.



Rys. 1.1-4 Główne rodzaje analityki: od opisu przeszłości do zautomatyzowanego podejmowania decyzji.

Pełnoprawna transformacja cyfrowa, która oznacza przejście do analityki systemowej i zarządzania opartego na danych, wymaga nie tylko outsourcingu, ale także utworzenia kompetentnego zespołu wewnętrznego. Kluczowymi członkami takiego zespołu powinni być menedżerowie produktu, inżynierowie danych, analitycy i programiści, którzy będą ściśle współpracować z jednostkami biznesowymi (rys. 4.3-9). Współpraca ta jest niezbędna do zadawania inteligentnych pytań analitycznych i skutecznego parametryzowania zadań związanych z podejmowaniem decyzji biznesowych. W społeczeństwie informacyjnym dane stają się nie tylko narzędziem pomocniczym, ale podstawą prognozowania i optymalizacji.

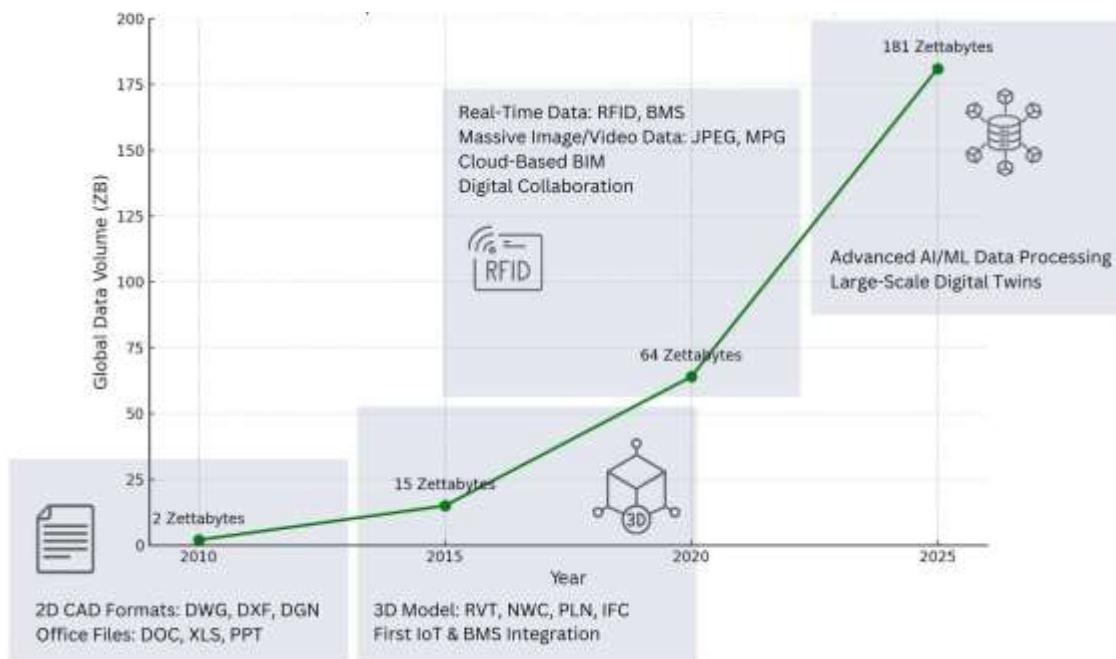
W budownictwie transformacja cyfrowa zasadniczo zmienia sposób projektowania, zarządzania i obsługi obiektów. Proces ten określany jest jako cyfryzacja informacji - gdzie wszystkie aspekty procesu budowlanego są digitalizowane do postaci cyfrowej nadającej się do analizy.

Cyfryzacja informacji o procesach budowlanych

Przez tysiąclecia ilość informacji rejestrowanych w budownictwie prawie się nie zmieniała, ale w ostatnich dziesięcioleciach gwałtownie wzrosła (Rys. 1.1-5).

Według badania PwC® "Managed Data. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9], 90% wszystkich danych na świecie powstało w ciągu ostatnich dwóch lat (stan na 2015 r.). Jednak większość firm nie wykorzystuje w pełni tych danych, ponieważ albo pozostają one w systemach silosowych, albo są po prostu archiwizowane bez rzeczywistej analizy.

Wzrost ilości danych przyspieszył tylko w ostatnich latach, podwajając się z 15 zettabajtów w 2015 roku do 181 zettabajtów w 2025 roku [10]. Każdego dnia serwery firm budowlanych i projektowych zapełniają się dokumentacją projektową, harmonogramami prac, obliczeniami i kalkulacjami, raportami finansowymi. Dla rysunków 2D/3D używane są formaty DWG, DXF i DGN, a dla modeli 3D - RVT, NWC, PLN i IFC™. Dokumenty tekstowe, tabele i prezentacje są zapisywane w formatach DOC, XLSX i PPT. Oprócz wideo i zdjęć z placu budowy - w MPG i JPEG, dane w czasie rzeczywistym z komponentów IoT, tagów RFID® (identyfikacja i śledzenie) oraz systemów zarządzania budynkiem BMS (monitorowanie i kontrola)



Rys. 1.1-5 Paraboliczny wzrost danych w latach 2010-2025 (na podstawie [10]).

Wraz z szybkim rozwojem informacji, branża budowlana staje przed koniecznością nie tylko gromadzenia i przechowywania danych, ale także zapewnienia ich weryfikacji, validacji, mieralności i przetwarzania analitycznego. Obecnie branża przechodzi aktywną fazę cyfryzacji informacji - systematycznego przekształcania wszystkich aspektów działalności budowlanej w formę cyfrową nadającą się do analizy, interpretacji i automatyzacji.

Cyfryzacja informacji oznacza pobieranie informacji o wszystkich podmiotach i elementach projektu budowlanego oraz samego procesu budowlanego - w tym tych, których wcześniej w ogóle nie uważaliśmy za informacje - i przekształcanie ich w format danych, aby informacje były wymierne i łatwe do analizy.

W kontekście budownictwa oznacza to przechwytywanie i digitalizację informacji na temat wszystkich elementów projektów i wszystkich procesów - od ruchu maszyn i ludzi na placu budowy po warunki pogodowe i klimatyczne na placu budowy, bieżące ceny materiałów i stopy procentowe banku centralnego - w celu generowania modeli analitycznych.

Jeśli potrafisz zmierzyć to, o czym mówisz i wyrazić to w liczbach, to wiesz coś na ten temat. Ale jeśli nie możesz wyrazić tego ilościowo, twoja wiedza jest bardzo ograniczona i niezadowalająca. Może to być punkt wyjścia, ale nie jest to poziom prawdziwej wiedzy naukowej. [11].

- W. Thomson (Lord Kelvin), 1824-1907, brytyjski naukowiec

Cyfryzacja informacji wykracza daleko poza tradycyjne podejście do gromadzenia informacji, w którym rejestrowano tylko podstawowe wskaźniki, takie jak roboczogodziny lub rzeczywiste koszty materiałów. Obecnie praktycznie każde zdarzenie można przekształcić w strumień danych nadający się do dogłębnej analizy przy użyciu zaawansowanych narzędzi analitycznych i technik uczenia maszynowego. Branża budowlana przeszła fundamentalne przejście od papierowych rysunków, arkuszy kalkulacyjnych Excel i ustnych instrukcji do systemów cyfrowych (Rys. 1.2-4), w których każdy element projektu staje się źródłem danych. Nawet pracownicy - od inżynierów po robotników budowlanych na placu budowy - są obecnie postrzegani jako zbiór cyfrowych zmiennych i zestawów danych.

Według raportu KPMG "Familiar Challenges - New Approaches: Global Construction Survey 2023", cyfrowe bliźniaki, sztuczna inteligencja (AI) i Big Data stają się kluczowymi czynnikami poprawiającymi rentowność projektów [2].

Nowoczesne technologie nie tylko upraszczają gromadzenie informacji, czyniąc je w dużej mierze automatycznym, ale także radykalnie obniżają koszty przechowywania danych. W rezultacie firmy odchodzą od selektywnego podejścia i wolą przechowywać cały szereg informacji do późniejszej analizy (rys. 2.1-5), co otwiera potencjalne możliwości optymalizacji procesów w przyszłości.

Cyfryzacja informacji i digitalizacja umożliwiają odkrycie ukrytej, wcześniej niewykorzystanej wartości informacji. Odpowiednio zorganizowane dane mogą być ponownie wykorzystane, zinterpretowane i zintegrowane z nowymi usługami i rozwiązaniami.

W przyszłości cyfryzacja informacji prawdopodobnie doprowadzi do pełnej automatyzacji zarządzania dokumentami, wprowadzenia samodzielnie zarządzanych procesów budowlanych i powstawania się nowych zawodów - analityków danych budowlanych, ekspertów ds. zarządzania projektami AI i inżynierów cyfrowych. Projekty budowlane staną się dynamicznymi źródłami informacji, a podejmowanie decyzji nie będzie oparte na intuicji lub subiektywnym doświadczeniu, ale na wiarygodnych i powtarzalnych faktach cyfrowych.

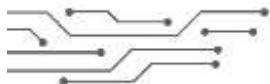
Informacje to ropa XXI wieku, a analityka to silnik spalinowy [12].

- Peter Sondergaard, starszy wiceprezes, Gartner®

Według IoT Analytics 2024 [13], oczekuje się, że globalne wydatki na zarządzanie danymi i analitykę wzrosną dramatycznie z 185,5 mld USD w 2023 r. do 513,3 mld USD do 2030 r., przy złożonej rocznej stopie wzrostu wynoszącej 16%. Jednak nie wszystkie komponenty rosną w tym samym tempie: analityka rośnie szybko, podczas gdy wzrost pamięci masowej spowalnia. Analityka zapewni najszybszy wzrost w ekosystemie zarządzania danymi: prognozuje się, że wzrośnie z 60,6 mld USD w 2023 r. do 227,9 mld USD do 2030 r., co oznacza złożoną roczną stopę wzrostu na poziomie 27%.

Wraz z przyspieszającą cyfryzacją informacji i szybkim wzrostem ich ilości, zarządzanie projektami budowlanymi i firmami staje przed koniecznością systematycznego przechowywania, analizowania i przetwarzania różnorodnych, często heterogenicznych danych. W odpowiedzi na to wyzwanie, począwszy od połowy lat 90-tych, branża rozpoczęła masowe przejście na elektroniczne tworzenie, przechowywanie i zarządzanie dokumentacją - od arkuszy kalkulacyjnych i obliczeń projektowych po rysunki i umowy.

Tradycyjne dokumenty papierowe, wymagające podpisów, fizycznego przechowywania, regularnej weryfikacji i archiwizacji w szafach, są stopniowo zastępowane przez systemy cyfrowe, które przechowują dane w ustrukturyzowany sposób - w bazach danych wyspecjalizowanych aplikacji.



ROZDZIAŁ 1.2.

TECHNOLOGIE I SYSTEMY ZARZĄDZANIA W NOWOCZESNYM BUDOWNICTWIE

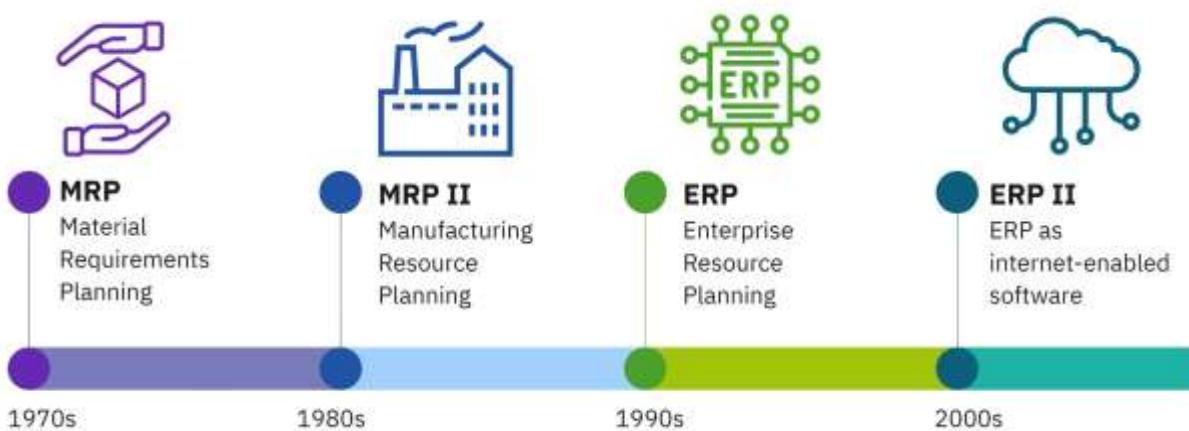
Rewolucja cyfrowa i pojawienie się modułowych systemów MRP/ERP -systems

Era nowoczesnego cyfrowego przechowywania i przetwarzania danych rozpoczęła się wraz z pojawieniem się taśmy magnetycznej w latach 50-tych XX wieku, która otworzyła możliwość przechowywania i wykorzystywania dużych ilości informacji. Kolejnym przełomem było pojawienie się napędów dyskowych, które radikalnie zmieniły podejście do zarządzania danymi w branży budowlanej.

Wraz z rozwojem hurtowni danych, wiele firm weszło na rynek rozwiązań i zaczęło opracowywać modułowe oprogramowanie do tworzenia, przechowywania, przetwarzania danych i automatyzacji rutynowych zadań

Gwałtowny wzrost ilości informacji i narzędzi doprowadził do zapotrzebowania na zintegrowane, modułowe rozwiązania, które nie pracują z pojedynczymi plikami, ale pomagają zarządzać i kontrolować przepływ danych w procesach i projektach.

Pierwsze kompleksowe narzędzia platformy musiały nie tylko przechowywać dokumenty, ale także dokumentować wszystkie żądania zmian i operacje w procesach: kto je zainicjował, jaki był zakres żądania i co ostatecznie zostało zapisane jako wartość lub atrybut. Do tych celów potrzebny był system, który mógłby śledzić dokładne obliczenia i podejmowane decyzje (rys. 1.2-1). Takimi platformami były pierwsze systemy MRP (Material Requirements Planning) i ERP (Enterprise Resource Planning), które stały się popularne na początku lat 90-tych [14]

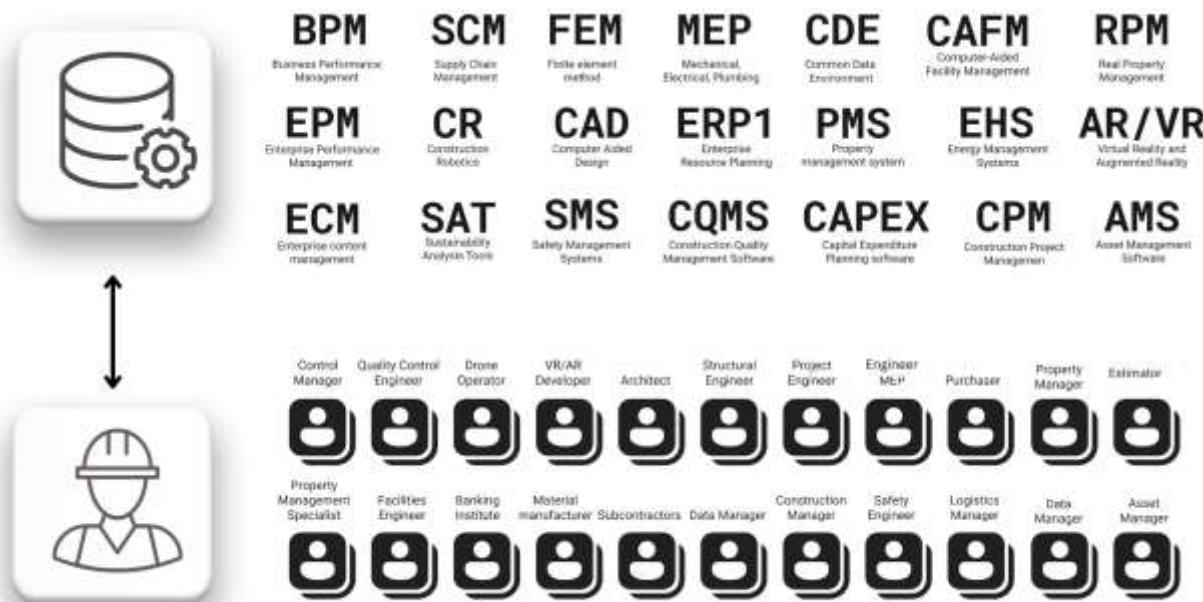


Rys. 1.2-1 Postęp w technologii przechowywania danych doprowadził do pojawienia się systemów ERP w latach 80-tych.

Pierwsze systemy MRP - i ERP - położyły podwaliny pod erę cyfryzacji w zarządzaniu procesami biznesowymi i projektami budowlanymi. Systemy modułowe, pierwotnie zaprojektowane w celu automatyzacji kluczowych procesów biznesowych, zostały ostatecznie zintegrowane z dodatkowymi,

bardziej elastycznymi i adaptacyjnymi rozwiązaniami programowymi.

Te dodatkowe rozwiązania zostały zaprojektowane do przetwarzania danych i zarządzania treścią projektu (rys. 1.2-2), albo zastąpiły niektóre moduły dużych systemów, albo skutecznie je uzupełniły, rozszerzając funkcjonalność całego systemu.



Rys. 1.2-2 Nowe rozwiązania programowe przyciągnęły do firmy armię menedżerów, którzy zarządzają przepływem danych.

W ciągu ostatnich dziesięcioleci firmy intensywnie inwestowały w systemy modułowe [15], postrzegając je jako długoterminowe, zintegrowane rozwiązania.

Według raportu Software Path na rok 2022 [16], średni budżet na użytkownika systemu ERP wynosi 9 000 USD. Średnio około 26% pracowników firmy korzysta z takich systemów. Zatem dla organizacji posiadającej 100 użytkowników całkowity koszt wdrożenia systemu ERP wynosi około 900 000 USD.

Inwestycje we własne, zamknięte, modułowe rozwiązania stają się coraz mniej uzasadnione w kontekście szybkiego rozwoju nowoczesnych, elastycznych i otwartych technologii. Jeśli takie inwestycje zostały już poczynione, ważne jest, aby obiektywnie ponownie ocenić rolę istniejących systemów: czy pozostają one niezbędne w dłuższej perspektywie, czy też ich funkcje można zmienić i wdrożyć w bardziej wydajny i przejrzysty sposób.

Jednym z kluczowych problemów dzisiejszych modułowych platform danych jest to, że centralizują one zarządzanie danymi w ramach zamkniętych aplikacji. W rezultacie dane - podstawowy zasób firmy - stają się zależne od konkretnych rozwiązań programowych, a nie odwrotnie. Ogranicza to ponowne wykorzystanie informacji, komplikuje migrację i zmniejsza elastyczność biznesową w szybko zmieniającym się krajobrazie cyfrowym.

Jeśli jest prawdopodobne, że istotność lub znaczenie zamkniętej architektury modułowej zmniejszy się w przyszłości, sensowne jest uznanie poniesionych dziś kosztów za koszty utopione i skupienie się na

strategicznym przejściu na bardziej otwarty, skalowalny i adaptacyjny ekosystem cyfrowy.

Oprogramowanie własnościowe charakteryzuje się wyłączną kontrolą dewelopera nad kodem źródłowym i danymi użytkownika tworzonymi w ramach korzystania z takich rozwiązań. W przeciwieństwie do oprogramowania open source, użytkownicy nie mają dostępu do wewnętrznej struktury aplikacji i nie mogą jej samodzielnie przeglądać, modyfikować czy dostosowywać do swoich potrzeb. Zamiast tego są oni zobowiązani do zakupu licencji, które przyznają prawo do korzystania z oprogramowania w granicach określonych przez sprzedawcę.

Nowoczesne podejście skoncentrowane na danych oferuje inny paradymat: dane powinny być postrzegane jako główny zasób strategiczny - niezależny, trwały i oddzielony od konkretnych rozwiązań programowych. Aplikacje z kolei stają się jedynie narzędziami do przetwarzania danych, które można swobodnie wymieniać bez ryzyka utraty krytycznych informacji.

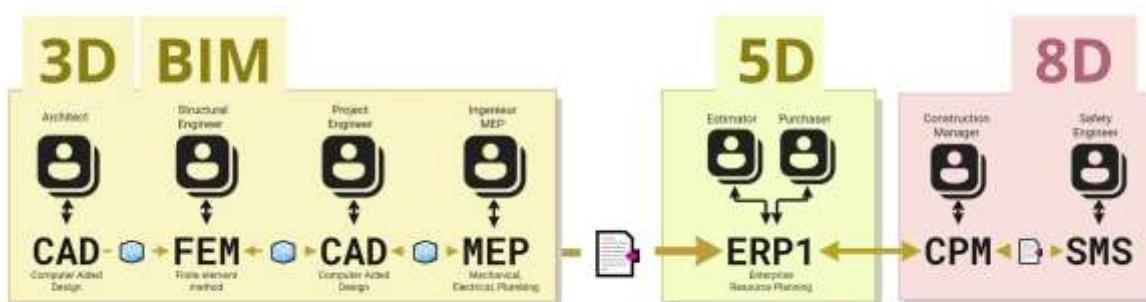
Rozwój systemów ERP i MRP w latach 90-tych (Rys. 1.2-1) zapewnił firmom potężne narzędzia do zarządzania procesami, ale miał również niezamierzoną konsekwencję w postaci znacznego zwiększenia liczby osób zaangażowanych w utrzymanie przepływu informacji. Zamiast automatyzować i upraszczać zadania operacyjne, systemy te często tworzyły nowe poziomy złożoności, biurokracji i zależności od wewnętrznych zasobów IT.

Systemy zarządzania danymi: od eksploracji danych do wyzwań biznesowych

Współczesne firmy stają przed koniecznością integracji wielu systemów zarządzania danymi. Wybór systemów zarządzania danymi, mądro zarządzanie tymi systemami i integracja różnych źródeł danych staje się krytyczna dla wydajności biznesowej.

W połowie 2020 roku można znaleźć setki (i tysiące w dużych firmach budowlanych) różnych systemów (Rys. 1.2-3), które muszą działać w harmonii, aby wszystkie aspekty procesu budowlanego przebiegały płynnie i spójnie.

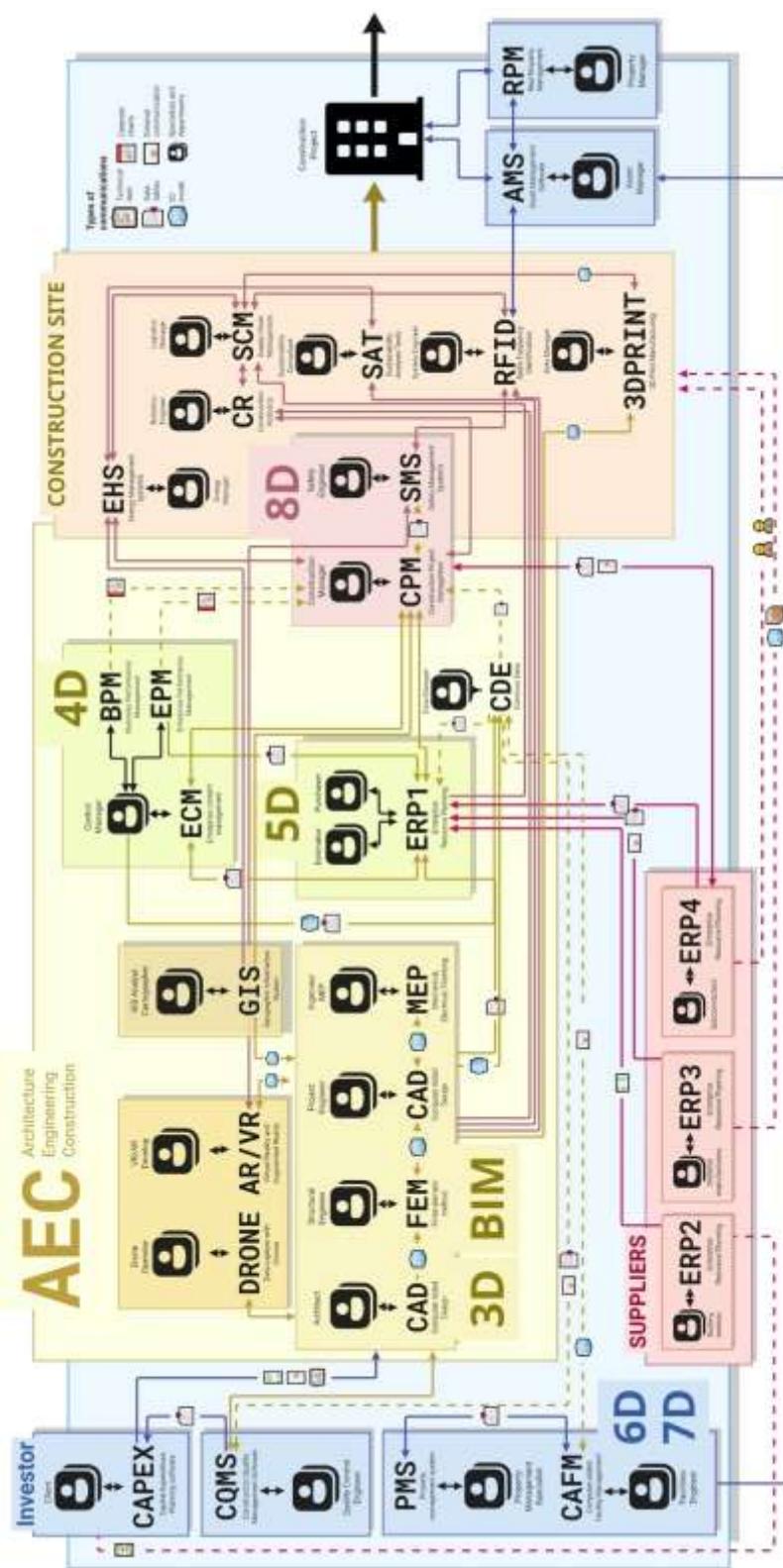
Według badania Deloitte z 2016 r.® "Data-Driven Management in Digital Capital Projects" - przeciętny specjalista budowlany korzysta z 3,3 aplikacji dziennie, ale tylko 1,7 z nich jest ze sobą zintegrowanych [17].



Rys. 1.2-3 Każdy system biznesowy wymaga profesjonalnego zespołu i odpowiedzialnego menedżera do zarządzania danymi wysokiej jakości.

Poniżej znajduje się lista popularnych systemów dla średnich i dużych firm z branży budowlanej, które są wykorzystywane do efektywnego zarządzania projektami budowlanymi:

- **ERP (Enterprise Resource Planning)** - zapewnia integrację procesów biznesowych, w tym księgowości, zaopatrzenia i zarządzania projektami.
- **CAPEX (Capital Expenditure Planning Software)** - służy do budżetowania i zarządzania inwestycjami finansowymi w projekty budowlane, pomaga określić koszt środków trwałych i inwestycji w aktywa długoterminowe.
- **CAD (Computer-Aided Design) i BIM (Building Information Modeling)** - służą do tworzenia szczegółowych i dokładnych rysunków technicznych oraz modeli 3D projektów. Systemy te koncentrują się na pracy z informacjami geometrycznymi.
- **MEP (Mechanical, Electrical, Plumbing)** - systemy inżynierijne obejmujące komponenty mechaniczne, elektryczne i hydrauliczne oraz szczegółowo opisujące wewnętrzny system "krążenia" projektu.
- **GIS (Systemy Informacji Geograficznej)** - wykorzystywane do analizy i planowania terenu, w tym kartografii i analizy przestrzennej.
- **CQMS (oprogramowanie do zarządzania jakością w budownictwie)** - zapewnia zgodność procesów budowlanych z ustalonymi normami i przepisami, pomagając wyeliminować usterki.
- **CPM (zarządzanie projektem budowlanym)** - obejmuje planowanie, koordynację i kontrolę procesów budowlanych.
- **CAFM (Computer-Aided Facility Management)** - systemy zarządzania i utrzymania budynków.
- **SCM (zarządzanie łańcuchem dostaw)** jest wymagane w celu optymalizacji przepływu materiałów i informacji między dostawcami a placem budowy.
- **EPM (Enterprise Performance Management)** - ma na celu poprawę procesów biznesowych i wydajności.
- **AMS (Asset Management Software)** - służy do optymalizacji użytkowania, zarządzania i konserwacji sprzętu i infrastruktury w całym cyklu życia aktywów.
- **RPM (Real Property Management)** - obejmuje zadania i procesy związane z zarządzaniem i eksploatacją budynków i gruntów, a także powiązanych zasobów i aktywów.



Rys. 1.2-4 Wzajemne powiązania między systemami, które łączą procesy firmy z przepływem informacji między różnymi działami.

- **CAE (Computer-Aided Engineering)** - Inżynieria wspomagana komputerowo, obejmuje systemy obliczeniowe i symulacyjne, takie jak analiza elementów skończonych (FEA) i obliczeniowa dynamika płynów (CFD).
- **CFD (Computational Fluid Dynamics)** - obliczeniowa dynamika płynów, modelowanie przepływów płynów i gazów. Podkategoria CAE.
- **CAPP (Computer-Aided Process Planning)** - komputerowo wspomagane planowanie procesów. Służy do tworzenia map tras i procesów.
- **CAM (Computer-Aided Manufacturing)** - komputerowe wspomaganie wytwarzania, generowanie programów sterujących dla maszyn CNC.
- **PDM (Product Data Management)** - Zarządzanie danymi produktu, system przechowywania i zarządzania dokumentacją techniczną.
- **MES (Manufacturing Execution System)** to system kontroli procesów produkcyjnych w czasie rzeczywistym.
- **PLM (Product Lifecycle Management)** - zarządzanie cyklem życia elementu projektu, integruje PDM, CAPP, CAM i inne systemy dla pełnej kontroli produktu od rozwoju do utylizacji.

Te i wiele innych systemów, obejmujących różnorodne rozwiązania programowe, stały się integralną częścią nowoczesnego przemysłu budowlanego (rys. 1.2-4). U podstaw takich systemów leżą specjalistyczne bazy danych z intuicyjnymi interfejsami, które zapewniają wydajne wprowadzanie, przetwarzanie i analizę informacji na wszystkich etapach projektowania i budowy. Integracja narzędzi cyfrowych ze sobą nie tylko pomaga zoptymalizować procesy pracy, ale także znaczco poprawia trafność podejmowanych decyzji, co ma pozytywny wpływ na czas i jakość realizacji projektu.

Ale w połowie przypadków nie ma integracji. Według statystyk tylko co druga aplikacja lub system jest zintegrowany z innymi rozwiązaniami [17]. Wskazuje to na ciągłą fragmentację środowiska cyfrowego i podkreśla potrzebę opracowania otwartych standardów i ujednolicionych interfejsów w celu zapewnienia kompleksowej wymiany informacji w ramach projektu budowlanego.

Jednym z głównych wyzwań w integracji dla nowoczesnych firm pozostaje wysoka złożoność systemów cyfrowych i wymagania dotyczące kompetencji użytkowników niezbędnych do skutecznego wyszukiwania i interpretacji informacji. Zespół specjalistów, kierowany przez kluczowego menedżera, jest tworzony w celu wsparcia każdego systemu wdrożonego w przedsiębiorstwie (rys. 1.2-2).

Kluczowy menedżer systemu odgrywa decydującą rolę we właściwym kierunku przepływu danych i jest odpowiedzialny za jakość ostatecznych informacji, podobnie jak pierwi menedżerowie tysiące lat temu byli odpowiedzialni za liczby zapisane na papirusie lub glinianych tabliczkach.

Aby przekształcić rozproszone przepływy informacji w narzędzie do zarządzania, niezbędna jest zdolność do systematycznej integracji danych i zarządzania nimi. W tej architekturze menedżerowie muszą działać jako elementy jednej sieci - jak grzybnia, która łączy poszczególne części firmy w spójny żywy organizm zdolny do adaptacji i ewolucji.

Korporacyjna grzybnia: jak dane łączą się z procesami biznesowymi

Proces integracji danych w aplikacjach i bazach danych opiera się na agregacji informacji z różnych źródeł,

w tym z różnych działów i specjalistów (Rys. 1.2-4). Specjaliści wyszukują odpowiednie dane, przetwarzają je i przesyłają do swoich systemów i aplikacji w celu dalszego wykorzystania.

Każdy system firmy, składający się z zestawu narzędzi, technologii i baz danych, jest drzewem wiedzy zakorzenionym w glebie danych historycznych i rosnącym, aby przynosić nowe owoce w postaci gotowych rozwiązań: dokumentów, obliczeń, tabel, wykresów i pulpitów nawigacyjnych (rys. 1.2-5). Systemy w firmie, podobnie jak drzewa w konkretnym lesie, współpracują i komunikują się ze sobą, tworząc złożony i dobrze zorganizowany system wspierany i zarządzany przez menedżerów-ekspertów.

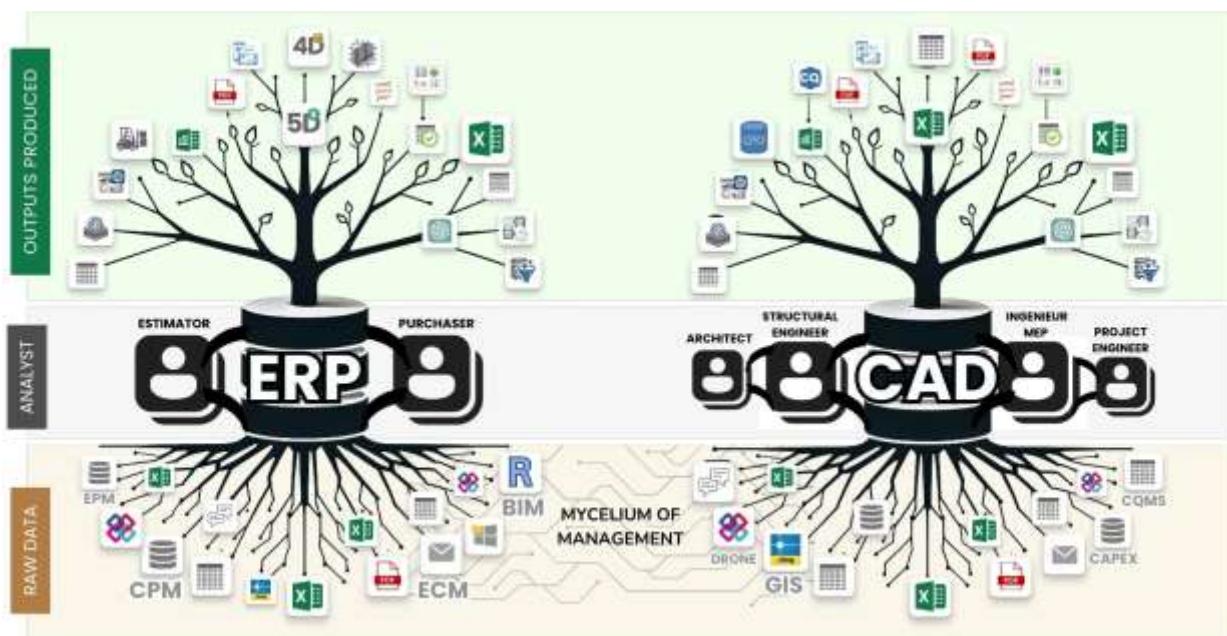
System wyszukiwania i przesyłania informacji w firmie działa jak złożona sieć leśna składająca się z drzew (systemów) i grzybni (menedżerów), które działają jako przewodnicy i recyklerzy, zapewniając, że informacje są przesyłane i przepływają do właściwych systemów. Pomaga to utrzymać zdrowy i wydajny przepływ i dystrybucję danych w firmie.

Eksperci, podobnie jak korzenie, absorbowią surowe dane na początkowych etapach projektu, przekształcając je w składniki odżywcze dla ekosystemu korporacyjnego. Systemy zarządzania danymi i treścią (rys. 1.2-4 - ERP, CPM, BIM, itp.) działają jak potężne autostrady informacyjne, przez które ta wiedza krąży na wszystkich poziomach firmy.

Podobnie jak w naturze, gdzie każdy element ekosystemu odgrywa swoją rolę, w krajobrazie biznesowym firmy każdy uczestnik procesu - od inżyniera do analityka - przyczynia się do wzrostu i płodności środowiska informacyjnego. Te systemowe "drzewa danych" (rys. 1.2-5) to nie tylko mechanizmy gromadzenia informacji, ale przewaga konkurencyjna, która zapewnia firmie zrównoważony rozwój.

Ekosystemy leśne są zaskakująco dokładnym odzwierciedleniem tego, jak zorganizowane są cyfrowe struktury korporacyjne. Podobnie jak wielopoziomowa struktura lasu - od zarośli po wierzchołki drzew - ład korporacyjny przypisuje zadania do poziomów odpowiedzialności i działów funkcjonalnych.

Głębokie i rozgałęzione korzenie drzew zapewniają odporność i dostęp do składników odżywczych. Podobnie, solidna struktura organizacyjna i stabilne procesy pracy z danymi wysokiej jakości wspierają cały ekosystem informacyjny firmy, przyczyniając się do jej zrównoważonego wzrostu i rozwoju nawet w okresach niestabilności rynku i kryzysów.



Rys. 1.2-5 Integracja danych za pośrednictwem różnych systemów jest jak grzybnia, która łączy menedżerów i specjalistów w jedną sieć informacyjną.

Współczesne rozumienie skali w biznesie ewoluowało. Dziś wartość firmy jest określana nie tylko przez jej widoczną część - "korony" w postaci dokumentów końcowych i raportów - ale także przez głębokość "systemu korzeniowego" jakościowo gromadzonych i systematycznie przetwarzanych danych. Im więcej informacji można zebrnąć i przetworzyć, tym wyższa staje się wartość biznesowa. Firmy, które metodycznie gromadzą "kompost" już przetworzonych danych i są w stanie wydobyć z niego przydatne spostrzeżenia, zyskują strategiczną przewagę

Informacje historyczne stają się nowym rodzajem kapitału, umożliwiającym wzrost, optymalizację procesów i przewagę konkurencyjną. W świecie opartym na danych wygrywa nie ten, kto ma ich więcej, ale ten, kto wie więcej.

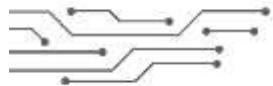
Dla branży budowlanej oznacza to przejście do zarządzania projektami w czasie rzeczywistym, gdzie wszystkie procesy - od projektowania i zaopatrzenia po koordynację wykonawców - będą oparte na odpowiednich, codziennie aktualizowanych danych. Integracja informacji pochodzących z różnych źródeł (systemy ERP, modele CAD, czujniki IoT na placach budowy, RFID) umożliwi tworzenie dokładniejszych prognoz, szybkie reagowanie na zmiany i unikanie opóźnień spowodowanych brakiem aktualnych danych.

Według McKinsey & Company's Data-Driven Enterprise 2025 (McKinsey & Company®, 2022 [18]), odnoszące sukcesy firmy przyszłości będą polegać na danych we wszystkich kluczowych aspektach swojej działalności, od decyzji strategicznych po interakcje operacyjne.

Dane przestaną być jedynie narzędziem analitycznym i staną się integralną częścią wszystkich procesów biznesowych, zapewniając przejrzystość, kontrolę i automatyzację zarządzania. Oparty na danych ruch pod adresem pozwoli organizacjom zminimalizować wpływ czynnika ludzkiego, zmniejszyć ryzyko operacyjne oraz zwiększyć przejrzystość i efektywność podejmowania decyzji.

XXI wiek wywraca paradygmat ekonomiczny do góry nogami: podczas gdy ropa naftowa była kiedyś

nazywana "czarnym złotem" ze względu na jej zdolność do napędzania maszyn i transportu, dziś, skompresowane pod presją czasu, dane historyczne stają się nowym strategicznym zasobem, zasilającym nie maszyny, ale algorytmy decyzyjne, które będą napędzać biznes.



ROZDZIAŁ 1.3.

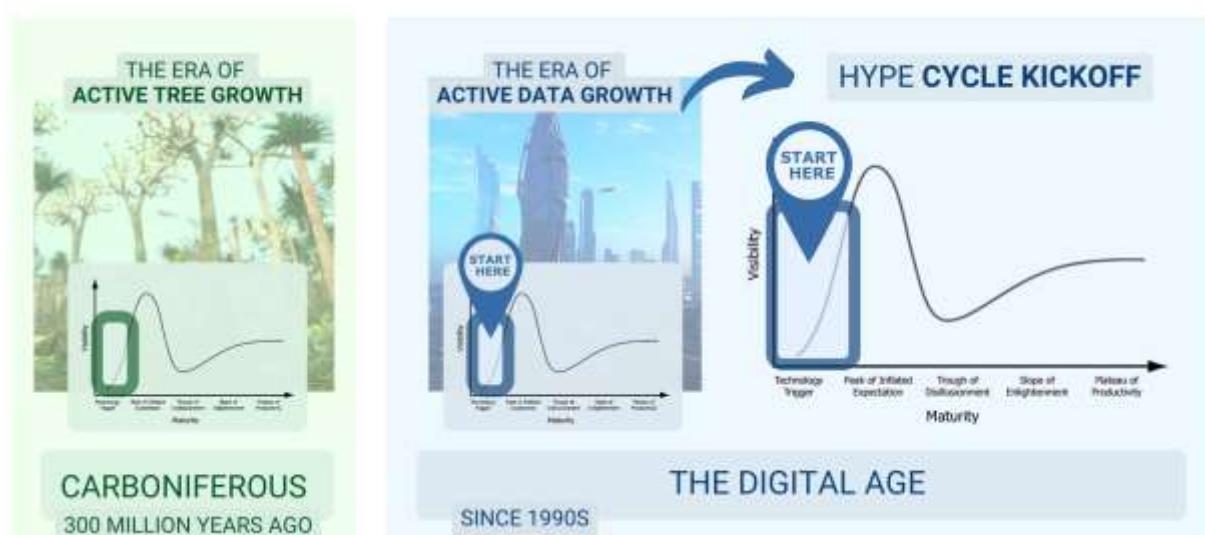
REWOLUCJA CYFROWA I EKSPLOZJA DANYCH

Początek boomu na dane jako fala ewolucyjna

Branża budowlana doświadcza bezprecedensowej eksplozji informacji. Jeśli pomyślimy o biznesie jako o drzewie wiedzy (rys. 1.2-5) zasilanym danymi, obecny etap cyfryzacji można porównać do gwałtownego wzrostu roślinności w okresie karbonu, epoki, w której biosfera Ziemi została przekształcona przez szybką akumulację biomasy (rys. 1.3-1).

Wraz z globalną cyfryzacją, ilość informacji w branży budowlanej podwaja się każdego roku. Nowoczesna technologia umożliwia gromadzenie danych w tle, analizowanie ich w czasie rzeczywistym i wykorzystywanie na skalę, która jeszcze niedawno wydawała się niemożliwa.

Zgodnie z prawem Moore'a, sformułowanym przez Gordona Moore'a (współzałożyciela firmy Intel®), gęstość i złożoność układów scalonych oraz ilość przetwarzanych i przechowywanych danych podwaja się mniej więcej co dwa lata [19].



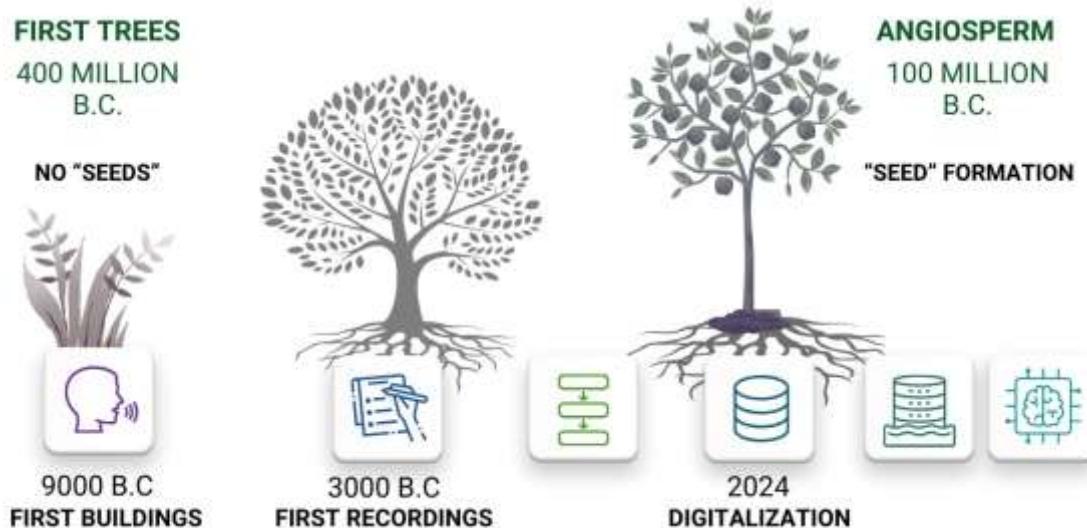
Rys. 1.3-1 Początek cyfryzacji doprowadził do wykładniczego wzrostu ilości danych, podobnie jak gwałtowny wzrost roślinności w erze węgla.

Podczas gdy starożytne megalityczne struktury, takie jak Göbekli Tepe (Turcja), nie pozostawiły po sobie udokumentowanej wiedzy nadającej się do ponownego wykorzystania, dzisiejsze technologie cyfrowe umożliwiają gromadzenie i ponowne wykorzystywanie informacji. Można to porównać do ewolucyjnego przejścia od roślin zarodnikowych do roślin nasiennych (okrytozałapkowych): pojawienie się nasion dało początek powszechnemu rozprzestrzenianiu się życia na naszej planecie. (Rys. 1.3-2).

Podobnie, dane z poprzednich projektów stają się rodzajem "cyfrowych nasion" - nośników wiedzy DNA, które mogą być skalowane i wykorzystywane w nowych projektach i produktach. Pojawienie się nowoczesnych narzędzi sztucznej inteligencji - uczenia maszynowego i dużych modeli językowych (LLM), takich jak ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok - umożliwia automatyczne

wyodrębnianie, interpretowanie i stosowanie danych w nowych kontekstach

Podobnie jak nasiona zrewolucjonizowały rozprzestrzenianie się życia na początkowo pozbawionej życia planecie, "nasiona danych" stają się podstawą do automatycznego pojawiania się nowych struktur informacyjnych i wiedzy, umożliwiając cyfrowym ekosystemom niezależną ewolucję i dostosowywanie się do zmieniających się wymagań użytkowników.



Rys. 1.3-2 Cyfrowe "nasiona danych" odgrywają taką samą rolę ewolucyjną jak okrytozałążkowe, rośliny kwitnące, które przekształciły ekosystem Ziemi.

Stoimy u progu nowej ery w budownictwie, w której eksplozja danych i aktywne rozpowszechnianie "nasion danych" - ustrukturyzowanych informacji z przeszłych i trwających projektów - stanowią podstawę cyfrowej przyszłości branży. Ich "zapylanie" za pomocą modeli językowych dużych zbiorów danych (LLM) pozwala nam nie tylko obserwować zmiany cyfrowe, ale także aktywnie uczestniczyć w tworzeniu samouczących się, adaptacyjnych ekosystemów. To nie jest ewolucja - to cyfrowa rewolucja, w której dane stają się głównym budulcem nowej rzeczywistości

Ilość danych w branży budowlanej dramatycznie rośnie ze względu na informacje z różnych dziedzin w całym cyklu życia projektów budowlanych. To ogromne nagromadzenie danych popchnęło branżę budowlaną w kierunku ery Big Data [20].

- Prof. Hang Yang, Wydział Inżynierii Lądowej i Architektury, Wuhan University of Technology, Wuhan,

Wzrost ilości danych w erze informacji przypomina procesy ewolucyjne w przyrodzie: tak jak rozwój lasów zmienił starożytny krajobraz planety, tak obecna eksplozja informacji zmienia krajobraz całej branży budowlanej.

Ilość danych generowanych w nowoczesnej firmie

W ciągu ostatnich dwóch lat utworzono 90% wszystkich istniejących danych na świecie [21]. Od 2023 roku każda osoba, w tym specjalisi z branży budowlanej, generuje około 1,7 megabajta danych na sekundę [22], a całkowita ilość danych na świecie osiągnie 64 zettabajty w 2023 roku i przewiduje się, że przekroczy 180 zettabajtów, czyli 180×10^{15} megabajtów, do 2025 roku [23].

Ta eksplozja informacji ma historyczny precedens - wynalezienie prasy drukarskiej przez Johanna Gutenberga w XV wieku. Zaledwie pięćdziesiąt lat po jej wprowadzeniu liczba książek w Europie podwoiła się: w ciągu kilku dekad wydrukowano tyle książek, ile zostało stworzonych ręcznie przez poprzednie 1200 lat [24]. Obecnie obserwujemy jeszcze szybszy wzrost: ilość danych na świecie podwaja się co trzy lata.

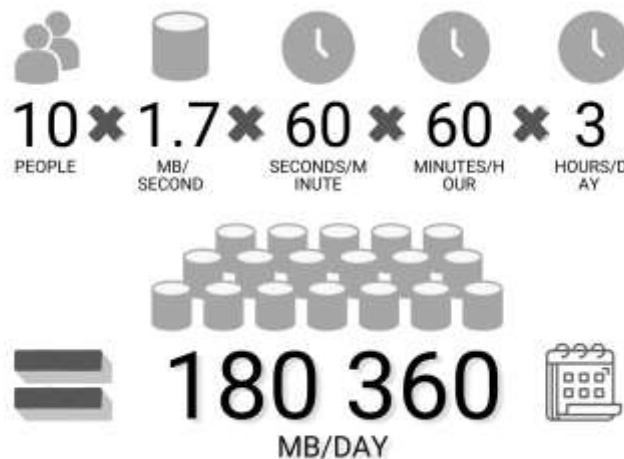
Biorąc pod uwagę obecne tempo przyrostu danych, branża budowlana może w ciągu najbliższych kilku dekad wygenerować tyle informacji, ile zgromadziła w całej swojej dotychczasowej historii



Rys. 1.3-3 Codzienne przechowywanie danych przez każdego pracownika na serwerach firmy przyczynia się do stałego wzrostu ilości danych.

W dzisiejszym świecie biznesu budowlanego nawet małe firmy generują codziennie ogromne ilości wieloformatowych informacji, a cyfrowy ślad nawet małej firmy budowlanej może sięgać dziesiątek gigabajtów dziennie - od modeli i rysunków po dokumentację fotograficzną i czujniki na placu budowy. Jeśli założymy, że każdy technik generuje średnio około 1,7 MB danych na sekundę, odpowiada to około 146 GB dziennie lub 53 TB rocznie (rys. 1.3-3).

Gdy zespół 10 osób pracuje aktywnie tylko przez 3 godziny dziennie, łączna ilość informacji generowanych w ciągu dnia sięga 180 gigabajtów (rys. 1.3-4).



Rys. 1.3-4 Firma zatrudniająca 10 osób generuje około 50-200 gigabajtów danych dziennie.

Zakładając, że 30% danych jest nowych (reszta jest nadpisywana lub usuwana), 10-osobowa firma może tworzyć kilkaset gigabajtów nowych danych miesięcznie (rzeczywiste liczby zależą od rodzaju działalności prowadzonej przez firmę).

Jest więc jasne: nie tylko generujemy coraz więcej danych - mamy do czynienia z rosnącą potrzebą ich efektywnego zarządzania, przechowywania i długoterminowej dostępności. I o ile wcześniej dane mogły "leżeć" na lokalnych serwerach bez żadnych kosztów, o tyle w kontekście cyfrowej transformacji coraz więcej firm zaczyna wykorzystywać rozwiązania chmurowe jako podstawę swojej infrastruktury informatycznej.

Koszt przechowywania danych: aspekt ekonomiczny

W ostatnich latach coraz więcej firm outsourcuje przechowywanie danych do usług w chmurze. Na przykład, jeśli firma przechowuje połowę swoich danych w chmurze, przy średniej cenie 0,015 USD za gigabajt miesięcznie, jej koszty przechowywania mogą wzrosnąć o 10-50 USD [25] każdego miesiąca.

Dla małej firmy z typowymi wzorcami generowania danych, koszty przechowywania danych w chmurze mogą ważyć się od setek do potencjalnie ponad tysiąca dolarów miesięcznie (Rys. 1.3-5) w ciągu kilku lat, tworząc potencjalnie znaczące obciążenie finansowe.

Według badania Forrester "Enterprises Outsource Data Storage as Complexity Grows" [26], w którym wzięło udział 214 osób podejmujących decyzje dotyczące infrastruktury technologicznej [26], w którym przebadano 214 osób podejmujących decyzje dotyczące infrastruktury technologicznej, ponad jedna trzecia organizacji korzysta z outsourcingu pamięci masowej, aby poradzić sobie z rosnącą ilością i złożonością operacji na danych, przy czym prawie dwie trzecie przedsiębiorstw preferuje model oparty na subskrypcji.



Rys. 1.3-5 Przeniesienie danych do chmury może zwiększyć miesięczne koszty przechowywania danych nawet o 2000 USD, nawet w przypadku firmy zatrudniającej zaledwie 10 pracowników.

Sytuację dodatkowo komplikuje przyspieszone wdrażanie technologii opartych na chmurze, takich jak CAD (BIM), CAFM, PMIS i ERP - systemów, które dodatkowo zwiększają koszty przechowywania i przetwarzania danych. W rezultacie firmy są zmuszone do poszukiwania sposobów optymalizacji kosztów i zmniejszenia zależności od dostawców usług w chmurze.

Od 2023 roku, wraz z aktywnym rozwojem dużych modeli językowych (LLM), podejście do przechowywania danych zaczęło się zmieniać. Coraz więcej firm myśli o odzyskaniu kontroli nad swoimi danymi, ponieważ przetwarzanie informacji na własnych serwerach staje się bezpieczniejsze i bardziej opłacalne.

W tym kontekście na pierwszy plan wysuwa się tendencja do odchodzenia od przechowywania i przetwarzania w chmurze tylko niezbędnych danych na rzecz lokalnego wdrażania korporacyjnych rozwiązań LLM i AI. Jak zauważył CEO Microsoftu w jednym z wywiadów [27], zamiast polegać na kilku oddzielnego aplikacji lub opartych na chmurze rozwiązaniami SaaS do wykonywania różnych zadań, agenci AI będą zarządzać procesami w bazach danych, automatyzując funkcje różnych systemów.

[...] stare podejście do tej kwestii [przetwarzania danych] polegało na tym, że jeśli pomyślisz o tym, jak różne aplikacje biznesowe radziły sobie z integracją, używały konektorów. Firmy sprzedawały licencje na te konektory i wokół tego ukształtował się model biznesowy. SAP [ERP] jest jednym z klasycznych przykładów: dostęp do danych SAP był możliwy tylko wtedy, gdy posiadało się odpowiedni konektor. Wydaje mi się więc, że coś podobnego pojawi się w przypadku [AI] interakcji z agentami [...]. Podejście, przynajmniej nasze, jest takie: myślę, że koncepcja istnienia aplikacji biznesowych prawdopodobnie upadnie w erze agentów [AI]. Ponieważ jeśli się nad tym zastanowić, są to zasadniczo bazy danych z logiką biznesową

- Satya Nadella, dyrektor generalny Microsoft, wywiad dla kanału BG2, 2024. [28]

W tym paradygmacie podejście LLM oparte na danych wykracza poza klasyczne systemy. Sztuczna inteligencja staje się pośrednikiem między użytkownikiem a danymi (rys. 2.2-3, rys. 2.2-4), eliminując

potrzebę stosowania wielu interfejsów pośredniczących i zwiększać wydajność procesów biznesowych. Więcej o tym podejściu do pracy z danymi powiemy w rozdziale "Przekształcanie chaosu w porządek i zmniejszanie złożoności".

Podczas gdy architektura przyszłości wciąż nabiera kształtów, firmy już teraz borykają się z konsekwencjami podjętych w przeszłości decyzji. Masowa cyfryzacja ostatnich dziesięcioleci, której towarzyszyło wprowadzenie rozproszonych systemów i niekontrolowane gromadzenie danych, doprowadziła do nowego problemu - przeciążenia informacyjnego.

Granice gromadzenia danych: od masy do znaczenia

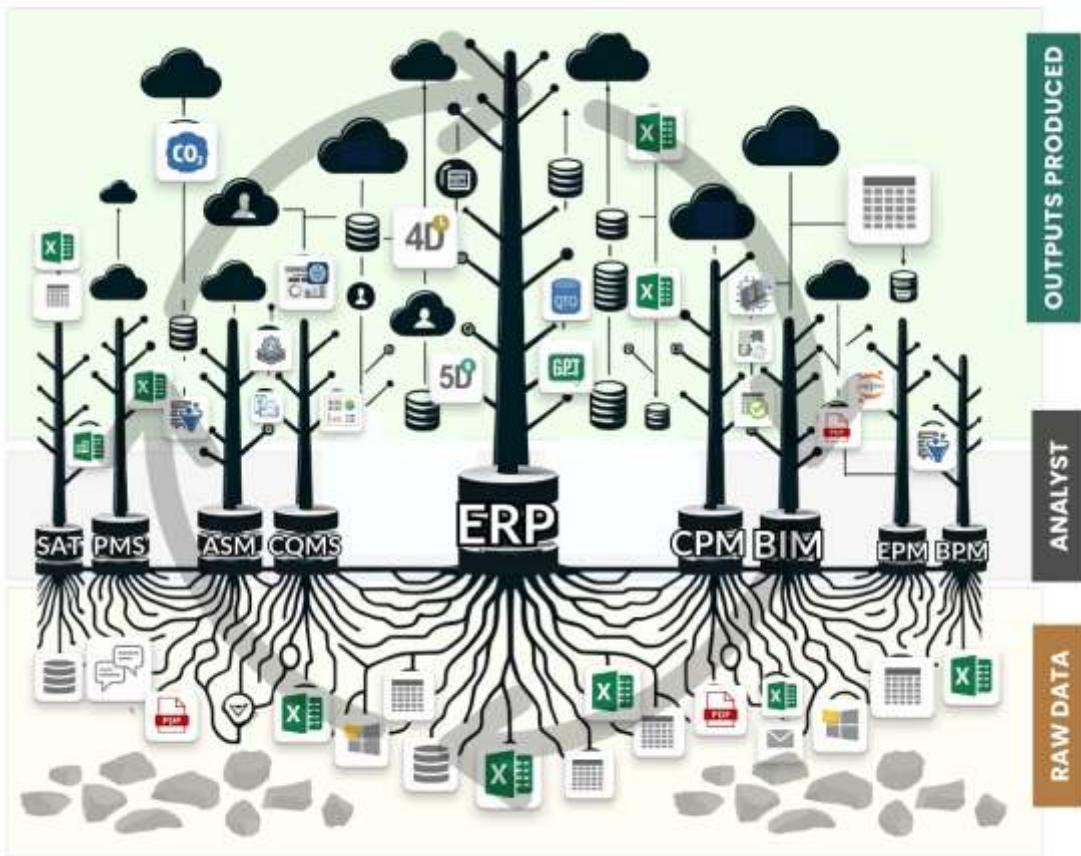
Nowoczesne systemy firmowe z powodzeniem rozwijają się i funkcjonują w warunkach kontrolowanego wzrostu, gdy ilość danych i liczba aplikacji są w równowadze z możliwościami działów IT i menedżerów. Jednak w ostatnich dziesięcioleciach cyfryzacja doprowadziła do niekontrolowanego wzrostu ilości i złożoności danych, co spowodowało efekt przesycenia w ekosystemie informacyjnym firm.

Obecnie serwery i pamięci masowe są narażone na bezprecedensowy napływ nieprzetworzonych i wieloformatowych informacji, które nie mają czasu zamienić się w kompost i szybko stają się nieistotne. Ograniczone zasoby firmy nie są w stanie poradzić sobie z tym zalewem, a dane gromadzą się w odizolowanych silosach (tzw. "silosach"), które wymagają ręcznego przetwarzania w celu wyodrębnienia przydatnych informacji.

W rezultacie, podobnie jak las porośnięty bluszczem i pokryty pleśnią, nowoczesne systemy zarządzania firmą często cierpią z powodu przeciążenia informacyjnego. Zamiast odżywczego humusu informacyjnego w sercu korporacyjnego ekosystemu, powstają odizolowane obszary danych w różnych formatach, co nieuchronnie prowadzi do spadku ogólnej wydajności procesów biznesowych.

Po długim okresie wykładniczego wzrostu ilości danych w ciągu ostatnich 40 lat nieuchronnie nastąpi faza nasycenia, a następnie ochłodzenia. Gdy pamięć masowa osiągnie swoje granice, nastąpi zmiana jakościowa: dane nie będą już tylko obiektem pamięci masowej, ale strategicznym zasobem.

Wraz z rozwojem sztucznej inteligencji i uczenia maszynowego, firmy mają możliwość obniżenia kosztów przetwarzania informacji i przejścia od ilościowego wzrostu do jakościowego wykorzystania danych. W ciągu następnej dekady branża budowlana będzie musiała przenieść swoją uwagę z tworzenia coraz większej ilości danych na zapewnienie ich struktury, integralności i wartości analitycznej.



Rys. 1.3-6 Odizolowane źródła danych uniemożliwiają wymianę informacji między systemami danych.

Główną wartością nie jest już ilość informacji, ale zdolność do ich automatycznej interpretacji i przekształcania w wiedzę użytkową przydatną do podejmowania decyzji zarządczych. Aby dane stały się naprawdę użyteczne, muszą być odpowiednio zarządzane: gromadzone, weryfikowane, strukturyzowane, przechowywane i analizowane w kontekście konkretnych zadań biznesowych.

Proces analizy danych w firmie jest podobny do cyklu życia i rozkładu drzew w lesie oraz pojawiania się nowych młodych i silnych drzew: dojrzałe drzewa umierają, rozkładają się i stają się pożywką dla nowego wzrostu. Ukończone i zakończone procesy, po ich zakończeniu, stają się częścią ekosystemu informacyjnego firmy, ostatecznie stając się humusem informacyjnym, który napędza przyszły rozwój nowych systemów i danych.

W praktyce jednak cykl ten jest często przerywany. Zamiast organicznej odnowy powstaje warstwowy chaos, podobny do warstw geologicznych, w którym nowe systemy są nakładane na stare bez głębokiej integracji i strukturyzacji. W rezultacie powstają rozproszone "silosy" informacyjne, utrudniające obieg wiedzy i komplikujące zarządzanie danymi.

Kolejne kroki: od teorii danych do praktycznych zmian

Ewolucja danych w budownictwie to podróz od glinianych tabliczek do nowoczesnych platform modułowych. Dzisiejszym wyzwaniem nie jest gromadzenie informacji, ale tworzenie ram, które przekształcają rozproszone i zróżnicowane dane w strategiczny zasób. Niezależnie od tego, czy jesteś kierownikiem firmy, czy inżynierem, zrozumienie wartości danych i sposobu pracy z nimi będzie kluczową umiejętnością w przyszłości.

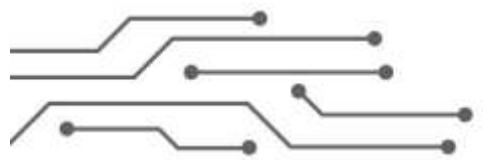
Podsumowując tę część, warto podkreślić główne praktyczne kroki, które pomogą ci zastosować omówione podejścia w codziennych zadaniach:

- Przeprowadzenie osobistego audytu przepływu informacji
 - Sporządź listę wszystkich systemów i aplikacji, z którymi pracujesz na co dzień.
 - Zaznacz miejsca, w których spędzasz najwięcej czasu na wyszukiwaniu lub podwójnym sprawdzaniu danych.
 - Identyfikacja kluczowych źródeł informacji
 - Analiza obecnego środowiska aplikacji pod kątem redundancji i powielania funkcji.
- Staraj się przechodzić przez procesy według poziomów dojrzałości analitycznej.
 - Rozpoczni swoje zadania od analizy opisowej (co się stało?).
 - Stopniowe wprowadzanie diagnostyki (dlaczego tak się stało?).
 - Zastanów się, w jaki sposób w procesach możesz przejść do analityki predykcyjnej (co się stanie?) i preskryptywnej (co należy zrobić?).
- Zacznij strukturyzować swoje dane robocze
 - Wdrożenie ujednoliconego systemu nazywania plików i folderów często używanych w pracy.
 - Tworzenie szablonów dla często używanych dokumentów i raportów
 - Regularne archiwizowanie ukończonych projektów o przejrzystej strukturze

Nawet jeśli nie możesz zmienić całej infrastruktury informacyjnej w swoim zespole lub firmie, zacznij od własnych procesów i drobnych usprawnień w codziennej pracy. Pamiętaj, że prawdziwą wartością danych nie jest ich ilość, ale zdolność do wyciągania z nich praktycznych wniosków. Nawet niewielkie, ale odpowiednio ustrukturyzowane i przeanalizowane zestawy danych mogą mieć znaczący wpływ, gdy zostaną zintegrowane z procesami decyzyjnymi.

W kolejnych częściach książki przejdziemy do konkretnych metod i narzędzi do pracy z danymi, przyjrzymy się sposobom przekształcania niestrukturyzowanych informacji w ustrukturyzowane zbiory, zbadamy technologie automatyzacji analityki i szczegółowo opiszemy, jak zbudować skuteczny ekosystem analityczny w firmie budowlanej.





ROZDZIAŁ 1.4.

JAK BRANŻA BUDOWLANA TONIE W CHAOSIE DANYCH

Druga część poświęcona jest krytycznej analizie wyzwań stojących przed firmami budowlanymi w radzeniu sobie z rosnącą ilością danych. Szczegółowo przeanalizowano konsekwencje fragmentacji informacji i zjawiska "danych w silosach", które utrudniają skuteczne podejmowanie decyzji. Badane są problemy związane z podejściem HiPPO (Highest Paid Person's Opinion) i jego wpływem na jakość decyzji zarządczych w projektach budowlanych. Oceniono wpływ dynamicznych procesów biznesowych i ich rosnącej złożoności na przepływ informacji i wydajność operacyjną. Podano konkretne przykłady tego, jak nadmierna złożoność systemu zwiększa koszty i zmniejsza elastyczność organizacji. Szczególną uwagę poświęcono ograniczeniom wynikającym z zastrzeżonych formatów oraz perspektywom wykorzystania otwartych standardów w branży budowlanej. Przedstawiono koncepcję przejścia na ekosystemy oprogramowania oparte na sztucznej inteligencji i LLM, które minimalizują nadmierną złożoność i bariery techniczne.

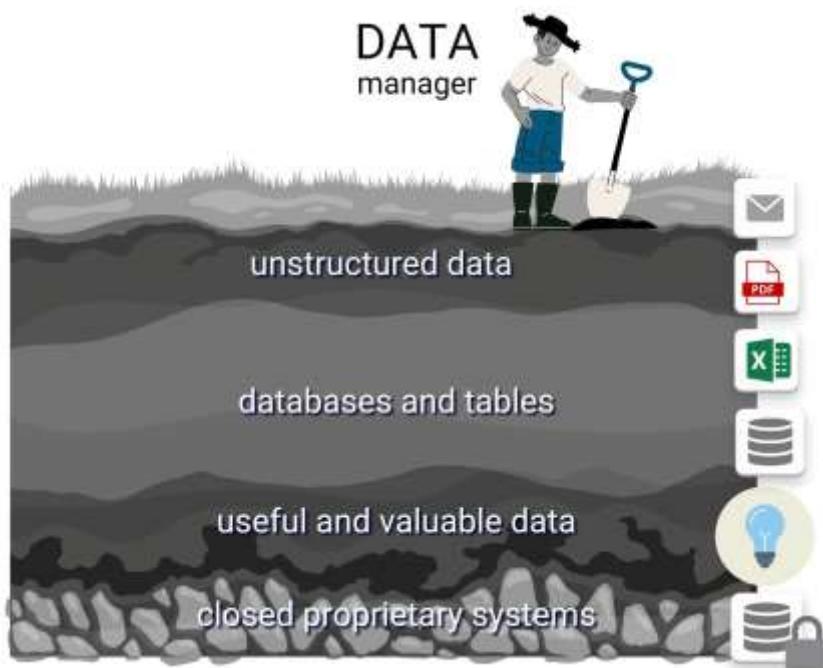
ROZDZIAŁ 1.5.

FRAGMENTACJA I SILOSY DANYCH

Im więcej narzędzi, tym bardziej wydajny biznes?

Na pierwszy rzut oka może się wydawać, że większa liczba narzędzi cyfrowych prowadzi dowiększej wydajności. W praktyce jednak tak nie jest. Z każdym nowym rozwiązaniem, czy to usługą w chmurze, starszym systemem czy kolejnym raportem Excel, firma dodaje kolejną warstwę do swojego cyfrowego krajobrazu - warstwę, która często nie jest zintegrowana z pozostałymi (Rys. 2.1-1).

Dane można porównać do węgla lub ropy naftowej: ich gromadzenie trwa latami, zagęszczając się pod warstwami chaosu, błędów, nieustrukturyzowanych procesów i zapomnianych formatów. Aby wydobyć z nich naprawdę przydatne informacje, firmy muszą dosłownie przedzierać się przez warstwy przestarzałych rozwiązań i cyfrowego szumu.



Rys. 2.1-1 Zróżnicowane dane tworzą podzielone warstwy - nawet "złote" spostrzeżenia giną w geologicznych skałach złożoności systemowej.

Każda nowa aplikacja pozostawia po sobie ślad: plik, tabelę lub cały odizolowany "silos" na serwerze. Jedna warstwa to glina (przestarzałe i zapomniane dane), druga to piasek (różne tabele i raporty), a trzecia to granit (zamknięte, zastrzeżone formaty, których nie można zintegrować). Z biegiem czasu środowisko cyfrowe firmy coraz bardziej przypomina zbiornik niekontrolowanej akumulacji informacji, w którym wartość jest tracona głęboko na serwerach firmy.

Z każdym nowym projektem i każdym nowym systemem nie tylko infrastruktura staje się bardziej złożona, ale także droga do użytecznych danych wysokiej jakości. Dotarcie do cennego "kamienia" wymaga

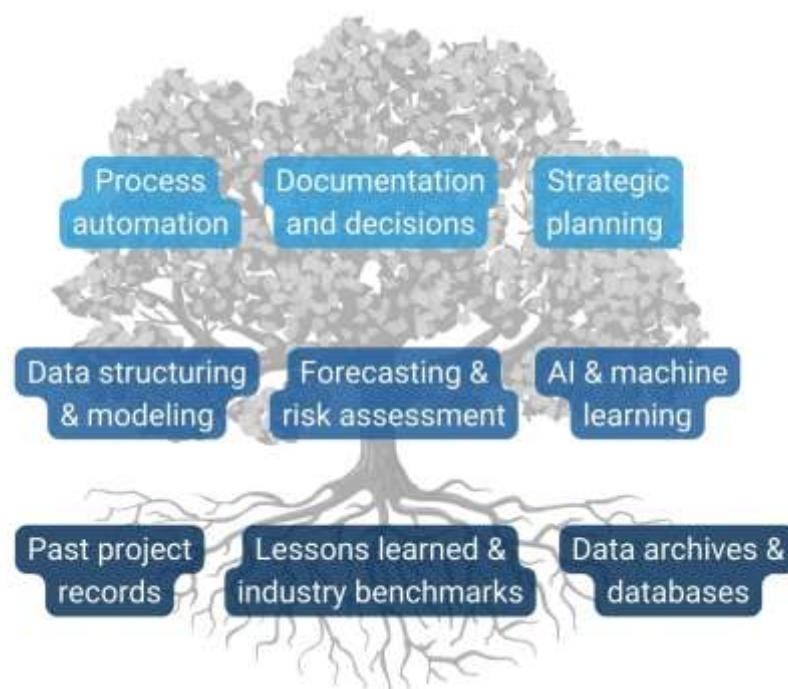
głębokiego oczyszczania, ustrukturyzowania informacji, pogrupowania ich w znaczące fragmenty i wyodrębnienia strategicznie ważnych spostrzeżeń poprzez analitykę i modelowanie danych.

Dane są cenną rzeczą i będą trwać dłużej niż same systemy [przetwarzające dane] [29].

- Tim Berners-Lee, ojciec sieci WWW i twórca pierwszej strony internetowej

Zanim dane staną się "wartościową rzeczą" i wiarygodną podstawą do podejmowania decyzji, muszą zostać starannie przygotowane. To właśnie odpowiednie przetwarzanie wstępne przekształca rozproszone dane w ustrukturyzowane doświadczenie, użyteczną prachnicę informacyjną, która następnie staje się narzędziem do prognozowania i optymalizacji.

Istnieje błędne przekonanie, że do rozpoczęcia analizy potrzebne są idealnie czyste dane, ale w praktyce możliwość pracy z brudnymi danymi jest istotną częścią procesu.



Rys. 2.1-2 Dane są systemem korzeniowym i fundamentem biznesu, który z kolei opiera się na procesach decyzyjnych.

Wraz z postępem technologicznym, Twoja firma musi również iść naprzód i nauczyć się tworzyć wartość z danych. Podobnie jak firmy naftowe i węglowe budują infrastrukturę do wydobywania minerałów, tak samo firmy muszą nauczyć się zarządzać przepływem nowych informacji na własnych serwerach i wydobywać cenne informacje z nieużywanych, niesformatowanych i nieaktualnych danych, przekształcając je w

strategiczny zasób.

Tworzenie pól (hurtowni danych) jest pierwszym krokiem. Nawet najpotężniejsze narzędzia nie rozwiążą problemu izolacji danych i danych wieloformatowych, jeśli firmy nadal będą działać w systemach silosowych. Gdy dane istnieją oddzielnie od siebie, bez przecinania się i udostępniania informacji, firmy stają w obliczu efektu "silosu danych". Zamiast jednej, spójnej infrastruktury, firmy są zmuszone do wydawania środków na łączenie i synchronizację danych.

Silosy danych i ich wpływ na wydajność firmy

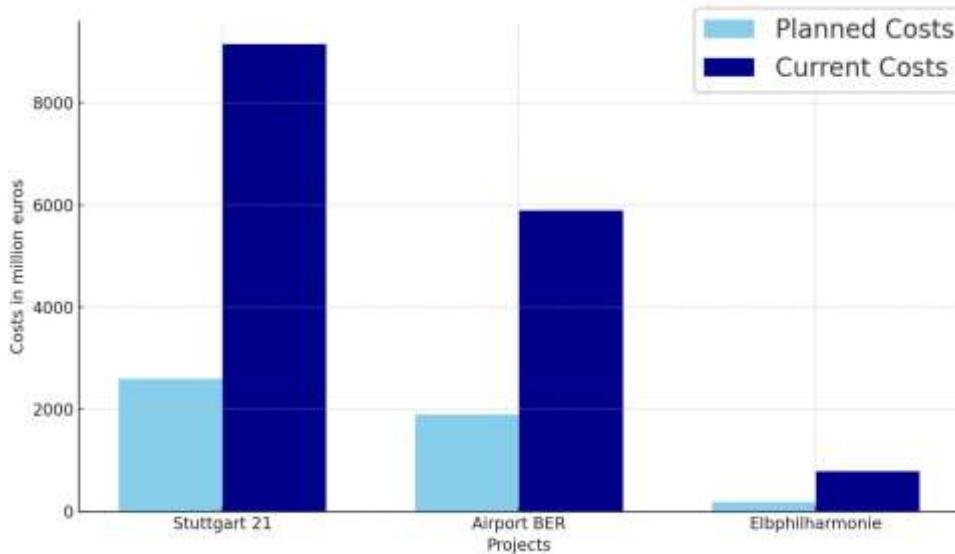
Wyobraź sobie, że budujesz osiedle mieszkaniowe, ale każdy zespół ma swój własny projekt. Niektórzy budują ściany, inni kładą linie komunikacyjne, a jeszcze inni kładą drogi, nie sprawdzając się nawzajem. W rezultacie rury nie pasują do otworów w ścianach, szyby windowe nie odpowiadają kondygnacjom, a drogi trzeba rozebrać i ułożyć ponownie.

Sytuacja ta nie jest tylko hipotetycznym scenariuszem, ale rzeczywistością wielu nowoczesnych projektów budowlanych. Ze względu na dużą liczbę generalnych wykonawców i podwykonawców pracujących z różnymi systemami i bez jednego centrum koordynującego, proces zamienia się w serię niekończących się zatwierrzeń, przeróbek i konfliktów. Wszystko to prowadzi do znacznych opóźnień i wzrostu kosztów projektu.

Klasyczna sytuacja na placu budowy jest prosta: szalunek jest gotowy, ale dostawa zbrojenia nie dotarła na czas. Sprawdzając informacje w różnych systemach, komunikacja wygląda mniej więcej następująco:

- ❷ 20. **brygadzista** na placu budowy pisze do kierownika projektu: "Skończyliśmy ustawać szalunki, gdzie są pręty zbrojeniowe?".
- ❷ **Kierownik projektu** (PMIS) do działu zaopatrzenia: - "Deskowanie jest gotowe. W moim systemie [PMIS] jest napisane, że pręty zbrojeniowe miały dotrzeć 18-go. Gdzie są pręty zbrojeniowe?"
- ❷ **Specjalista ds. łańcucha dostaw** (ERP): - "Nasz system ERP mówi, że dostawa będzie miała miejsce 25-go".
- ❷ **Inżynier danych** lub dział IT (odpowiedzialny za integracje): - W PMIS data to 18, w ERP to 25. Nie ma powiązania OrderID między ERP i PMIS, więc dane nie są zsynchronizowane. Jest to typowy przykład luki informacyjnej.
- ❷ **Kierownik projektu do dyrektora generalnego** - "Dostawa armatury jest opóźniona, plac budowy stoi, a kto jest za to odpowiedzialny jest niejasny".

Przyczyną incydentu była izolacja danych w różnych systemach. Integrując i ujednolicając źródła danych, tworząc jedno repozytorium informacji i automatyzując za pomocą narzędzi ETL (Apache NiFi, Airflow lub n8n), można wyeliminować silosy między systemami. Te i inne metody oraz narzędzia zostaną szczegółowo omówione w dalszej części książki.



Rys. 2.1-3 Porównanie planowanych i rzeczywistych kosztów dużych projektów infrastrukturalnych w Niemczech.

To samo dzieje się z systemami korporacyjnymi: najpierw tworzone są odizolowane rozwiązania, a następnie trzeba wydawać ogromne budżety na ich integrację i harmonizację. Gdyby modele danych i komunikacji były przemyślane od samego początku, integracja nie byłaby w ogóle potrzebna. Silosowe dane tworzą chaos w cyfrowym świecie, niczym nieskoordynowany proces budowlany.

Według badania KPMG z 2023 r. "Cue construction 4.0: Time to make or break", tylko 36% firm skutecznie udostępnia dane między działami, a 61% boryka się z poważnymi problemami z powodu odizolowanych "silosów" danych [30].



Rys. 2.1-4 Lata trudnych do zebrania danych gromadzą się w odizolowanych "silosach" pamięci masowej, ryzykując, że nigdy nie zostaną wykorzystane.

Dane firmowe są przechowywane w odizolowanych systemach, jak pojedyncze drzewa rozrzucone po

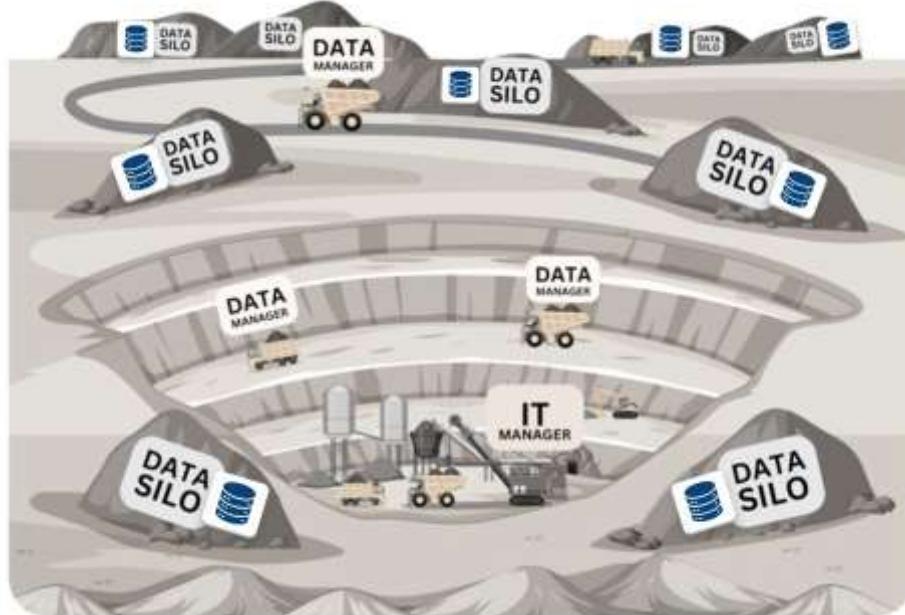
całym krajobrazie. Każde z nich zawiera cenne informacje, ale brak połączeń między nimi uniemożliwia stworzenie jednego, wzajemnie połączonego ekosystemu. Ta silosowość utrudnia przepływ danych i ogranicza zdolność organizacji do zobaczenia pełnego obrazu. Łączenie tych silosów jest niezwykle długim i złożonym procesem uprawy grzybni na poziomie zarządzania, aby dowiedzieć się, jak przenosić poszczególne fragmenty informacji między systemami.

Według badania WEF z 2016 r., jedną z głównych barier transformacji cyfrowej jest brak wspólnych standardów danych i fragmentacja.

Branża budowlana jest jedną z najbardziej rozdrobnionych na świecie i zależy od płynnej interakcji wszystkich uczestników łańcucha wartości [5].

- Światowe Forum Ekonomiczne 2016: Kształtowanie przyszłości

Projektanci, menedżerowie, koordynatorzy i programiści często wolą pracować autonomicznie, unikając złożoności koordynacji. Ta naturalna skłonność prowadzi do tworzenia "silosów" informacyjnych, w których dane są izolowane w oddzielnych systemach. Im więcej jest takich odizolowanych systemów, tym trudniej jest zmusić je do współpracy. Z czasem każdy system otrzymuje własną bazę danych i wyspecjalizowany dział wsparcia menedżerów (Rys. 1.2-4), co dodatkowo komplikuje integrację.



Rys. 2.1-5 Każdy system ma tendencję do tworzenia własnego unikalnego silosu danych, które muszą być przetwarzane przez odpowiednie narzędzia [31].

Błędne koło w systemach korporacyjnych wygląda następująco: firmy inwestują w złożone rozwiązania isolated, a następnie ponoszą wysokie koszty ich integracji, a programiści, zdając sobie sprawę ze złożoności łączenia systemów, wolą pracować w swoich zamkniętych ekosystemach. Wszystko to zwiększa fragmentację krajobrazu IT i utrudnia migrację do nowych rozwiązań (Rys. 2.1-5). Menedżerowie krytykują silosy danych, ale rzadko analizują ich przyczyny i sposoby zapobiegania im. Menedżerowie narzekają na

przestarzałe systemy IT, ale ich wymiana wymaga znacznych inwestycji i rzadko przynosi oczekiwane rezultaty. W rezultacie nawet próby walki z problemem często pogarszają sytuację.

Głównym powodem tego rozdżwięku jest przedkładanie aplikacji nad dane. Firmy najpierw opracowują oddzielne systemy lub kupują gotowe rozwiązania od dostawców, a następnie próbują je ujednolicić, tworząc zduplikowane i niekompatybilne pamięci masowe i bazy danych.

Przezwyciężenie problemu fragmentacji wymaga radykalnie nowego podejścia - przedkładania danych nad aplikacje. Firmy muszą najpierw opracować strategie zarządzania danymi i modele danych, a następnie zbudować systemy lub zakupić rozwiązania, które działają z jednym zestawem informacji, zamiast tworzyć nowe bariery.

Wkraczamy w nowy świat, w którym dane mogą być ważniejsze niż oprogramowanie.

- Tim O'Reilly, dyrektor generalny O'Reilly Media, Inc.

Badanie McKinsey Global Institute "Rethinking Construction: the path to improved productivity" (2016) pokazuje, że branża budowlana pozostaje w tyle za innymi sektorami w zakresie transformacji cyfrowej [32]. Według raportu, przyjęcie zautomatyzowanego zarządzania danymi i platform cyfrowych może znacząco poprawić produktywność i zmniejszyć straty związane z niespójnością procesów. Potrzeba transformacji cyfrowej została również podkreślona w raporcie Egana (Wielka Brytania, 1998) [33], który podkreśla kluczową rolę zintegrowanych procesów i podejścia opartego na współpracy w budownictwie.

W rezultacie, podczas gdy w ciągu ostatnich 10 000 lat głównym problemem dla zarządzających danymi był brak danych, wraz z lawiną danych i systemów zarządzania danymi, użytkownicy i menedżerowie stają w obliczu problemu - nadmiaru danych, co utrudnia znalezienie poprawnych pod względem prawnym i wysokiej jakości informacji.

Rozproszone silosy danych nieuchronnie prowadzą do poważnego problemu obniżonej jakości danych. W przypadku wielu niezależnych systemów te same dane mogą istnieć w różnych wersjach, często ze sprzecznymi wartościami, co stwarza dodatkowe trudności dla użytkowników, którzy muszą określić, które informacje są istotne i wiarygodne.

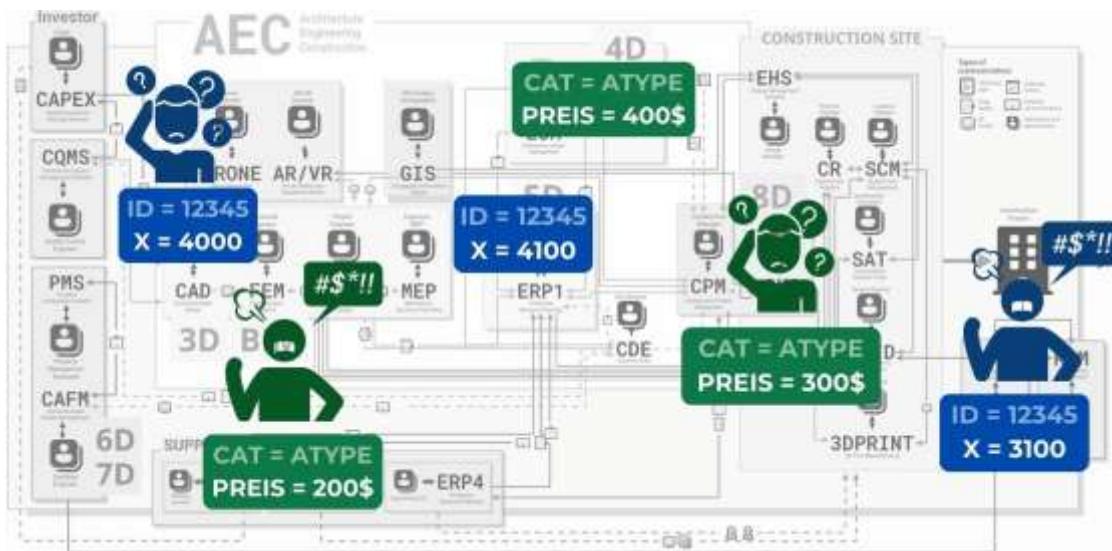
Powielanie i brak jakości danych jako konsekwencja braku jedności

Ze względu na problem silosów danych, menedżerowie muszą poświęcać dużo czasu na wyszukiwanie i uzgadnianie danych. Aby zabezpieczyć się przed problemami z jakością, firmy tworzą złożone struktury zarządzania informacjami, w których pion menedżerów jest odpowiedzialny za wyszukiwanie, weryfikację i uzgadnianie danych. Takie podejście zwiększa jednak biurokrację i spowalnia podejmowanie decyzji. Im więcej danych, tym trudniej je analizować i interpretować, zwłaszcza jeśli nie ma jednolitego standardu ich przechowywania i przetwarzania.

Wraz z mnogością aplikacji i systemów, które w ostatniej dekadzie wyrastały jak grzyby po deszczu, problem silosów i niewłaściwej jakości danych stał się coraz bardziej istotny dla użytkowników końcowych.

Te same dane, ale z różnymi wartościami, można teraz znaleźć w różnych systemach i aplikacjach (Rys. 2.1-6). Prowadzi to do trudności dla użytkowników końcowych, gdy próbują określić, która wersja danych jest odpowiednia i poprawna spośród wielu dostępnych. Prowadzi to do błędów w analizach i ostatecznie w podejmowaniu decyzji.

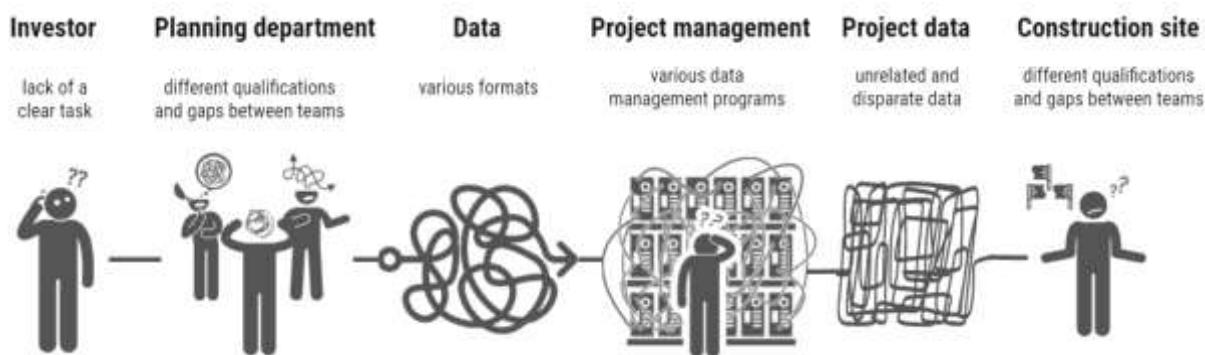
Aby zabezpieczyć się przed problemami ze znalezieniem właściwych danych, menedżerowie firm tworzą wielopoziomową biurokrację menedżerów ds. weryfikacji. Ich zadaniem jest szybkie znalezienie, sprawdzenie i przesłanie wymaganych danych w formie tabel i raportów, poruszając się w labiryncie różnych systemów.



Rys. 2.1-6 Próbując znaleźć właściwe dane, menedżerowie muszą zapewnić jakość i wiarygodność prawną danych między różnymi systemami.

W praktyce jednak model ten generuje nowe komplikacje. Gdy dane są zarządzane ręcznie, a informacje są rozproszone w wielu niepowiązanych decyzjach, każda próba uzyskania dokładnych i aktualnych informacji poprzez piramidę decydentów (Rys. 2.1-7) staje się wąskim gardłem - czasochłonnym i podatnym na błędy.

Sytuację pogarsza lawina rozwiązań cyfrowych. Rynek oprogramowania wciąż zalewany jest nowymi narzędziami, które wydają się obiecujące. Jednak bez jasnej strategii zarządzania danymi, rozwiązania te nie integrują się w jednolity system, a zamiast tego tworzą dodatkowe warstwy złożoności i powielania. W rezultacie, zamiast upraszczać procesy, firmy znajdują się w jeszcze bardziej rozdrobnionym i chaotycznym środowisku informacyjnym.



Rys. 2.1-7 Złożoność systemów i różnorodność formatów danych prowadzi do utraty spójności w procesie budowy.

Wszystkie te problemy związane z zarządzaniem wieloma różnymi rozwiązaniami przedzej czy później doprowadzają kierownictwo firmy do ważnego uświadomienia sobie: nie chodzi o ilość danych lub poszukiwanie kolejnego "uniwersalnego" narzędzia do przetwarzania danych. Prawdziwy powód leży w jakości danych i sposobie, w jaki organizacja je tworzy, otrzymuje, przechowuje i wykorzystuje.

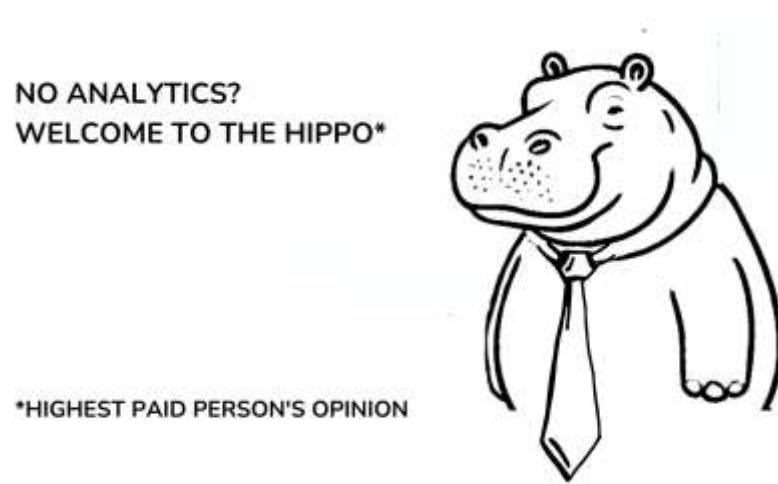
Kluczem do trwałego sukcesu nie jest pogoń za nowymi "magicznymi" aplikacjami, ale budowanie kultury danych w firmie. Oznacza to traktowanie danych jako strategicznego zasobu i uczynienie jakości, integralności i istotności danych priorytetem na wszystkich poziomach organizacji.

Rozwiązywanie dylematu "jakość kontra ilość" polega na stworzeniu ujednoliconej struktury danych, która eliminuje powielanie, niespójności i ujednolica przepływ informacji. Taka architektura zapewnia pojedyncze, niezawodne źródło danych, na podstawie których można podejmować świadome, dokładne i terminowe decyzje.

W przeciwnym razie, co nadal często ma miejsce, firmy nadal polegają na subiektywnych opiniach i intuicyjnych ocenach ekspertów HiPPO, a nie na wiarygodnych faktach. Jest to szczególnie widoczne w branży budowlanej, gdzie wiedza specjalistyczna tradycyjnie odgrywa znaczącą rolę.

HiPPO lub niebezpieczeństwo opinii w podejmowaniu decyzji

Tradycyjnie w branży budowlanej kluczowe decyzje podejmowane są w oparciu o doświadczenie i subiektywną ocenę. Bez aktualnych i wiarygodnych danych menedżerowie firmy muszą działać na ślepo, opierając się na intuicji najlepiej opłacanych pracowników (HiPPO - Highest Paid Person's Opinion), a nie na obiektywnych faktach (rys. 2.1-8).



Rys. 2.1-8 W przypadku braku analityki biznes zależy od subiektywnej opinii doświadczonych specjalistów.

Takie podejście może być uzasadnione w stabilnym i wolno zmieniającym się środowisku, ale w erze cyfrowej transformacji staje się poważnym zagrożeniem. Decyzje oparte na intuicji i domysłach są podatne na zniekształcenia, często opierają się na niepopartych hipotezach i nie uwzględniają złożonego obrazu odzwierciedlonego w danych

To, co uchodzi za inteligentną debatę na poziomie decyzyjnym w firmie, często nie jest oparte na niczym konkretnym. Sukces firmy nie powinien zależeć od autorytetu i wynagrodzenia ekspertów, ale od umiejętności efektywnej pracy z danymi, identyfikowania wzorców i podejmowania świadomych decyzji.

Ważne jest, aby porzucić przekonanie, że autorytet lub doświadczenie automatycznie oznaczają, że decyzja jest słuszna. Podejście oparte na danych zmienia zasady gry: dane i analityka, a nie stanowisko i wynagrodzenie, są teraz podstawą do podejmowania decyzji. Duże zbiory danych, uczenie maszynowe i analityka wizualna pozwalają nam identyfikować wzorce i opierać się na faktach, a nie na domysłach (rys. 1.1-4).

Bez danych jesteś tylko kolejną osobą z opinią [34].

- W. Edwards Deming, naukowiec i konsultant ds. zarządzania

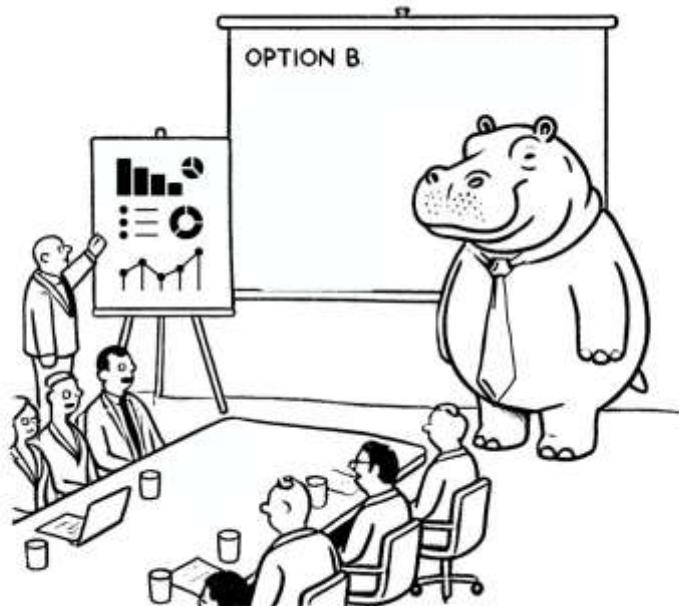
Nowoczesne metody zarządzania danymi zapewniają również ciągłość wiedzy w firmie. Jasno opisane procesy, automatyzacja i systematyczne podejście umożliwiają przeniesienie nawet kluczowych ról bez utraty efektywności.

Jednak ślepe zaufanie do danych może również prowadzić do poważnych błędów. Dane same w sobie są tylko zbiorem liczb. Bez odpowiedniej analizy, kontekstu i zdolności do identyfikowania wzorców, nie mają one żadnej wartości i nie mogą napędzać procesów. Klucz do sukcesu leży nie w wyborze między intuicją HiPPO a analityką, ale w budowaniu inteligentnych narzędzi, które przekształcają rozproszone informacje w możliwe do zarządzania, świadome decyzje.

W cyfrowym środowisku budowlanym to nie staż i miejsce w hierarchii stają się decydującymi czynnikami sukcesu, ale szybkość reakcji, trafność decyzji i wydajność zasobów.

Dane to narzędzia, a nie prawdy absolutne. Powinny one uzupełniać ludzkie myślenie, a nie je zastępować. Pomimo korzyści płynących z analityki, dane nie mogą całkowicie zastąpić ludzkiej intuicji i doświadczenia. Ich rolą jest pomoc w podejmowaniu bardziej trafnych i świadomych decyzji.

Przewaga konkurencyjna zostanie osiągnięta nie tylko dzięki spełnianiu standardów, ale także dzięki możliwości prześcignięcia konkurentów w efektywnym wykorzystaniu zasobów, które są takie same dla wszystkich. W przyszłości umiejętności związane z danymi staną się tak samo ważne, jak kiedyś umiejętności czytania i pisania czy umiejętności matematyczne. Specjalisi, którzy potrafią analizować i interpretować dane, będą w stanie podejmować trafniejsze decyzje, wypierając tych, którzy polegają wyłącznie na osobistym doświadczeniu (Rys. 2.1-9).



Rys. 2.1-9 Decyzje powinny być oparte na obiektywnej analizie, a nie na opinii najlepiej opłacanego pracownika.

Menedżerowie, specjaliści i inżynierowie będą działać jako analitycy danych, badając strukturę, dynamikę i kluczowe wskaźniki projektów. Zasoby ludzkie staną się elementami systemu, wymagającymi elastycznego dostosowania opartego na danych w celu maksymalizacji wydajności.

Błędy przy korzystaniu z nieodpowiednich danych są znacznie mniejsze niż w przypadku braku danych [35].

- Charles Babbage, wynalazca pierwszej analitycznej maszyny liczącej

Pojawienie się dużych zbiorów danych i wprowadzenie LLM (Large Language Models) radykalnie zmieniło nie tylko sposób analizy, ale także samą naturę podejmowania decyzji. Podczas gdy wcześniej skupiano się na przyczynowości (dlaczego coś się wydarzyło - analityka diagnostyczna) (rys. 1.1-4), dziś na pierwszy plan wysuwa się zdolność przewidywania przyszłości (analityka predykcyjna), a w przyszłości analityka preskryptywna, w której uczenie maszynowe i sztuczna inteligencja sugerują najlepszy wybór w procesie decyzyjnym.

Według nowego badania SAP™ "New Study Finds Nearly Half of Executives Trust Artificial Intelligence More Than Themselves" z 2025 roku [36], 44% kadry kierowniczej wyższego szczebla byłoby skłonnych zmienić swoją poprzednią decyzję w oparciu o porady AI, a 38% zaufałoby AI w podejmowaniu decyzji biznesowych w ich imieniu. Tymczasem 74% kadry kierowniczej stwierdziło, że ufa poradom sztucznej inteligencji bardziej niż swoim przyjaciołom i rodzinie, a 55% pracuje w firmach, w których spostrzeżenia oparte na sztucznej inteligencji zastępują lub często pomijają tradycyjne metody podejmowania decyzji - zwłaszcza w organizacjach o rocznych przychodach powyżej 5 miliardów dolarów. Ponadto 48% respondentów korzysta z narzędzi generatywnej sztucznej inteligencji na co dzień, w tym 15% używa ich wiele razy dziennie.

Wraz z rozwojem LLM i zautomatyzowanych systemów zarządzania danymi pojawia się nowe wyzwanie: jak efektywnie wykorzystywać informacje bez utraty ich wartości w chaosie niekompatybilnych formatów i heterogenicznych źródeł, który jest uzupełniany przez rosnącą złożoność i dynamikę procesów biznesowych.

Ciągły wzrost złożoności i dynamiki procesów biznesowych

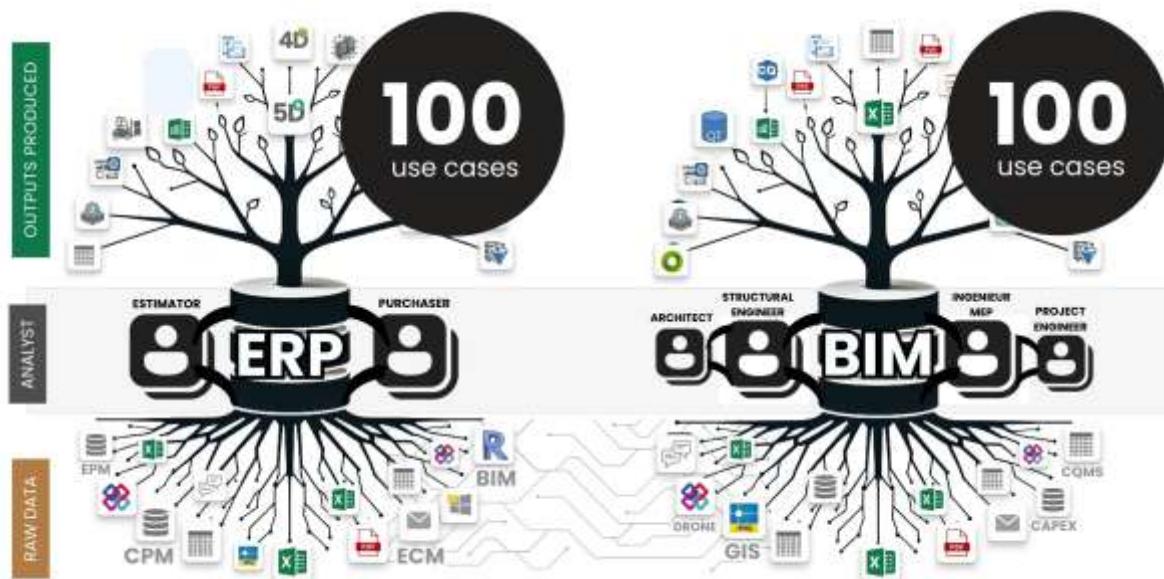
Branża budowlana stoi dziś przed poważnymi wyzwaniami w zakresie zarządzania danymi i procesami. Głównymi wyzwaniami są rozdrobnione systemy informatyczne, nadmierna biurokracja i brak integracji między narzędziami cyfrowymi. Wyzwania te nasilają się, ponieważ same procesy biznesowe stają się coraz bardziej złożone - napędzane przez technologię, zmieniające się wymagania klientów i ewoluujące przepisy.

Wyjątkowość projektów budowlanych wynika nie tylko z ich specyfiki technicznej, ale także z różnic w normach krajowych i wymogach regulacyjnych w różnych krajach (rys. 4.2-10, rys. 5.1-7). Wymaga to elastycznego, zindywidualizowanego podejścia do każdego projektu, co jest trudne do wdrożenia w ramach tradycyjnych modułowych systemów sterowania. Ze względu na złożoność procesów i dużą ilość danych, wiele firm zwraca się do dostawców oferujących specjalistyczne rozwiązania. Rynek jest jednak przeładowany - wiele startupów oferuje podobne produkty, skupiając się na wąskich zadaniach. W rezultacie holistyczne podejście do zarządzania danymi często traci na znaczeniu.

Adaptacja do ciągłego napływu nowych technologii i wymagań rynkowych staje się krytycznym czynnikiem konkurencyjności. Jednak istniejące aplikacje i systemy modułowe mają niską zdolność adaptacji - wszelkie zmiany często wymagają długich i kosztownych poprawek ze strony programistów, którzy nie zawsze rozumieją specyfikę procesów budowlanych.

Firmy stają się zakładnikami opóźnień technologicznych, czekając na nowe aktualizacje zamiast szybko

wdrażać innowacyjne, zintegrowane podejścia. W rezultacie wewnętrzna struktura organizacji budowlanych jest często złożonym ekosystemem wzajemnie powiązanych hierarchicznych i często zamkniętych systemów koordynowanych przez wielopoziomową sieć menedżerów (Rys. 2.1-10).

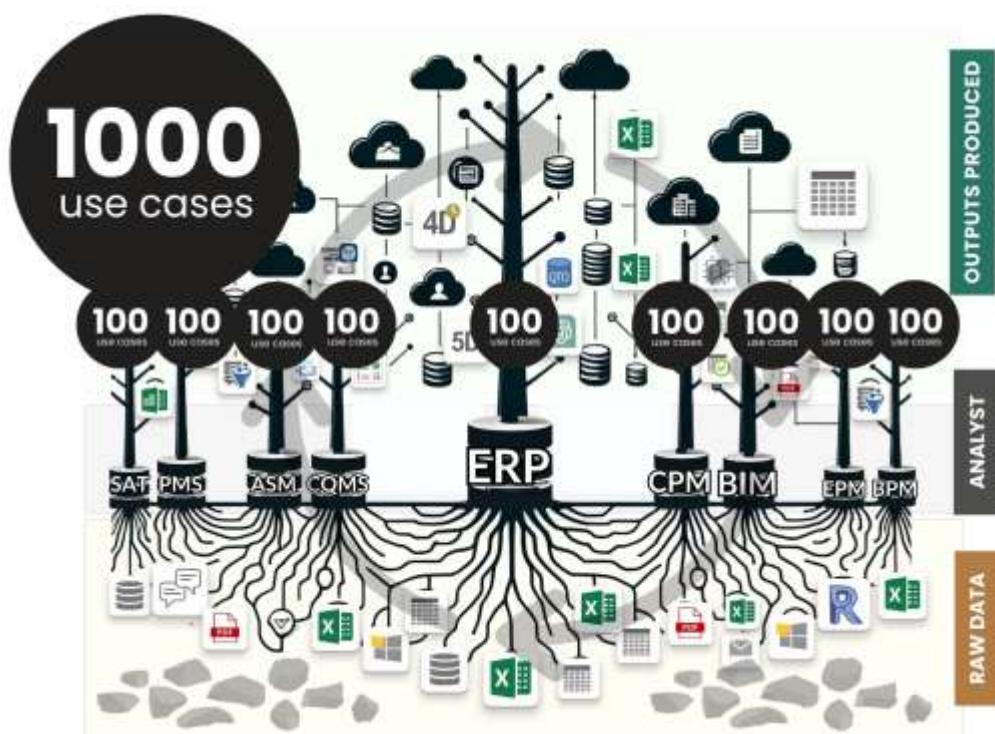


Rys. 2.1-10 Firmy składają się z połączonych ze sobą systemów, których wzajemne połączenia tworzą procesy wymagające automatyzacji.

Według badania przeprowadzonego przez Canadian Construction Association i KPMG Canada 2021 [37], tylko 25% firm uważa, że zajmuje znaczącą lub odmienną pozycję w porównaniu z konkurencją pod względem wdrażania technologii lub rozwiązań cyfrowych. Tylko 23% respondentów stwierdziło, że ich rozwiązania są w znacznym lub dużym stopniu oparte na danych. Jednocześnie większość respondentów badania określiła swoje wykorzystanie szeregu innych technologii jako czysto eksperymentalne lub przyznała, że w ogóle z nich nie korzysta.

Ta niechęć do udziału w eksperymentach technologicznych jest szczególnie widoczna w przypadku dużych projektów infrastrukturalnych, gdzie błędy mogą kosztować miliony dolarów. Nawet najbardziej zaawansowane technologie - cyfrowe bliźniaki, analityka predykcyjna - często spotykają się z oporem nie ze względu na ich skuteczność, ale z powodu braku udowodnionej niezawodności w rzeczywistych projektach.

Według raportu Światowego Forum Ekonomicznego (WEF) "Shaping the Future of Construction" [5], wprowadzenie nowych technologii w budownictwie napotyka nie tylko trudności techniczne, ale także bariery psychologiczne ze strony klientów. [5], wprowadzanie nowych technologii w budownictwie napotyka nie tylko trudności techniczne, ale także bariery psychologiczne ze strony klientów. Wielu klientów obawia się, że zastosowanie zaawansowanych rozwiązań uczyni z ich projektów poligon doświadczalny i sprawi, że staną się "królikami doświadczalnymi", a nieprzewidywalne konsekwencje mogą prowadzić do dodatkowych kosztów i ryzyka.



Rys. 2.1-11 Dla każdego przypadku użycia danych rynek rozwiązań oferuje aplikacje do optymalizacji i automatyzacji procesów.

Branża budowlana jest bardzo zróżnicowana: różne projekty mają różne wymagania, specyfikę regionalną, ustawowe przepisy klasyfikacyjne (rys. 4.2-10), normy obliczeniowe (rys. 5.1-7) itp. Dlatego praktycznie niemożliwe jest stworzenie własnej, uniwersalnej aplikacji lub systemu, który idealnie pasowałby do wszystkich tych wymagań i specyfiki projektu.

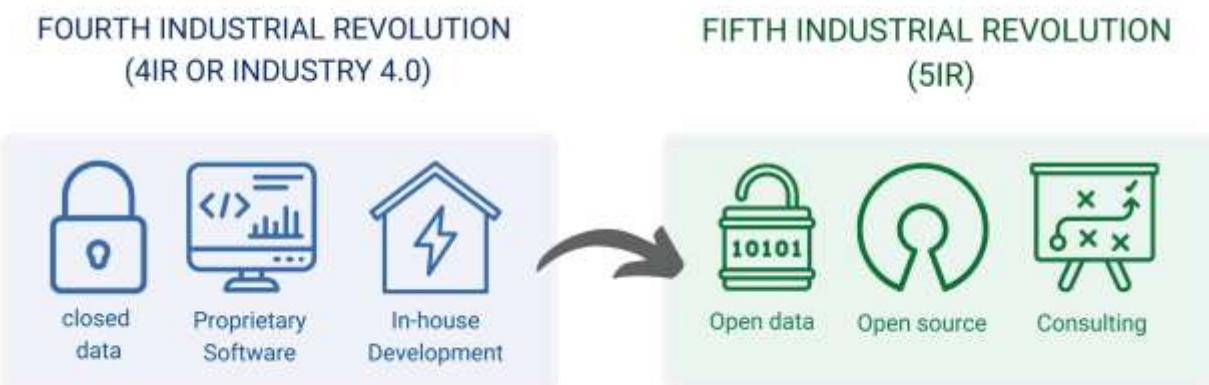
Próbując poradzić sobie z rosnącą złożonością systemów i zależnością od dostawców oprogramowania, coraz częściej zdaje się sobie sprawę, że kluczem do skutecznego zarządzania danymi jest nie tylko otwartość i standaryzacja, ale także uproszczenie samej architektury procesów. Rosnąca złożoność i dynamika procesów biznesowych wymaga nowego podejścia, w którym priorytet przesuwa się z gromadzenia danych na ich strukturyzację i organizację. To właśnie ta zmiana będzie kolejnym krokiem w rozwoju branży budowlanej, oznaczającym koniec ery dominacji dostawców oprogramowania i początek ery sensownej organizacji informacji.

Uświadomienie sobie ograniczeń uniwersalnych rozwiązań i podatności na rosnącą złożoność prowadzi do zmiany priorytetów z zamkniętych platform i gromadzenia danych na przejrzystość, zdolność adaptacji i uporządkowaną obsługę informacji. Ta zmiana sposobu myślenia odzwierciedla szersze zmiany w globalnej gospodarce i technologii, opisywane przez pryzmat tak zwanych "rewolucji przemysłowych". Aby zrozumieć, dokąd zmierza budownictwo i jaki jest jego przyszły kierunek, konieczne jest rozważenie miejsca branży w kontekście czwartej i piątej rewolucji przemysłowej - od automatyzacji i cyfryzacji po personalizację, otwarte standardy i model danych oparty na usługach.

Czwarta rewolucja przemysłowa (Przemysł 4.0) i piąta rewolucja przemysłowa (Przemysł 5.0) w budownictwie

Etapy technologiczne i ekonomiczne to koncepcje teoretyczne wykorzystywane do opisu i analizy ewolucji społeczeństwa i gospodarki na różnych etapach rozwoju. Mogą być one różnie interpretowane przez różnych badaczy i ekspertów.

- **Czwarta rewolucja przemysłowa** (4IR lub Przemysł 4.0) związana jest z technologią informacyjną, automatyzacją, cyfryzacją i globalizacją. Jednym z jej kluczowych elementów jest tworzenie własnych rozwiązań programowych, tj. wyspecjalizowanych produktów cyfrowych przeznaczonych do konkretnych zadań i firm. Rozwiązania te często stają się ważną częścią infrastruktury IT, ale są słabo skalowalne bez dodatkowych modyfikacji.
- **Piąta rewolucja przemysłowa (5IR)** znajduje się obecnie na wcześniejszym etapie konceptualizacji i rozwoju niż 4IR. Jej podstawowe zasady obejmują zwiększoną personalizację produktów i usług. 5IR to ruch w kierunku bardziej adaptowej, elastycznej i spersonalizowanej działalności gospodarczej, z naciskiem na personalizację, doradztwo i modele zorientowane na usługi. Kluczowym aspektem piątego trybu gospodarczego jest wykorzystanie danych do podejmowania decyzji, co jest praktycznie niemożliwe bez wykorzystania otwartych danych i otwartych narzędzi (Rys. 2.1-12).



Rys. 2.1-12 Czwarty wzorzec koncentruje się na rozwiązańach, podczas gdy piąty wzorzec koncentruje się na personalizacji i danych.

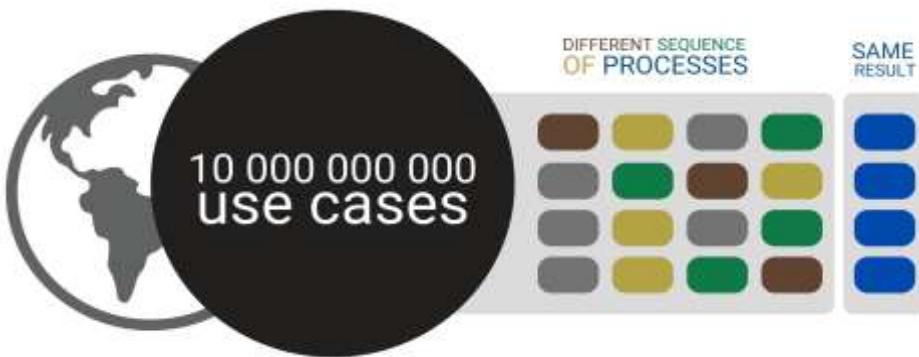
Stworzenie aplikacji dla firm z branży budowlanej do użytku w dziesięciu lub stu organizacjach nie gwarantuje jej skutecznego skalowania do innych firm, regionów lub krajów bez znaczących modyfikacji i ulepszeń. Prawdopodobieństwo skutecznego skalowania takich rozwiązań pozostaje niskie, ponieważ każda organizacja ma unikalne procesy, wymagania i warunki, które mogą wymagać spersonalizowanych dostosowań.

Ważne jest, aby zrozumieć, że już dziś udana integracja rozwiązań technologicznych oznacza głęboko spersonalizowane podejście do każdego procesu, projektu i firmy. Oznacza to, że nawet po opracowaniu uniwersalnej struktury, narzędzia lub programu, będzie on wymagał szczególowej adaptacji i dostosowania, aby spełnić unikalne wymagania i warunki każdej konkretnej firmy i projektu.

Według raportu PwC "Decoding the Fifth Industrial Revolution" [38] [38], około 50% kadry kierowniczej wyższego szczebla w różnych branżach polega w tym roku na integracji zaawansowanej technologii i ludzkiej wiedzy specjalistycznej. Takie podejście pozwala im szybko dostosowywać się do zmian w projekcie produktu lub wymaganiach klientów, tworząc spersonalizowaną produkcję.

Każdy proces wymaga opracowania unikalnej funkcji lub aplikacji, co, biorąc pod uwagę wielkość globalnej branży budowlanej i różnorodność projektów, prowadzi do istnienia ogromnej liczby przypadków biznesowych, z których każdy reprezentuje unikalną logikę Pipeline (rys. 2.1-13). Każdy taki przypadek ma swoją specyfikę i wymaga indywidualnego podejścia. Bardziej szczegółowo przyjrzymy się różnorodności możliwych rozwiązań tego samego problemu analitycznego w kontekście różnych podejść w rozdziale poświęconym uczeniu maszynowemu i parsowaniu zbioru danych Titanic (rys. 9.2-9).

Pipeline w kontekście procesów cyfrowych to sekwencja działań, procesów i narzędzi, które umożliwiają zautomatyzowany lub ustrukturyzowany przepływ danych i pracy przez różne etapy cyklu życia projektu.



Rys. 2.1-13 Indywidualność i zmienność przypadków biznesowych uniemożliwia próby stworzenia skalowalnych, zamkniętych platform i narzędzi.

Nasze życie zmieniło się już na wiele sposobów pod wpływem transformacji cyfrowej, a dziś możemy mówić o nadaniu nowego etapu w rozwoju gospodarczym branży budowlanej. W tej "nowej gospodarce" konkurencja będzie zorganizowana według innych zasad: ten, kto jest w stanie skutecznie przekształcić wiedzę publiczną i otwarte dane w pożądane produkty i usługi, zyskuje kluczową przewagę w warunkach piątej rewolucji przemysłowej.

Jak zauważa ekonomistka Kate Maskus w książce "Private Rights and Public Problems: The Global Intellectual Property Economy in the 21st Century" [39]. 2012 [39], "żyjemy w globalnej gospodarce opartej na wiedzy, a przyszłość należy do tych, którzy wiedzą, jak przekształcić odkrycia naukowe w towary".

Przejście do piątego trybu ekonomicznego oznacza przesunięcie punktu ciężkości z zamkniętych rozwiązań IT na otwarte standardy i platformy. Firmy zaczną odchodzić od tradycyjnego oprogramowania na rzecz modeli zorientowanych na usługi, w których głównym zasobem będą dane, a nie zastrzeżone technologie.

Badanie Harvard Business School 2024 [40] pokazuje ogromną wartość ekonomiczną oprogramowania open source (Open Source Software, OSS). Według badania, OSS jest obecny w 96% wszystkich kodów oprogramowania, a niektóre komercyjne oprogramowanie składa się w 99,9% z komponentów OSS. Bez OSS firmy wydawałyby 3,5 razy więcej na oprogramowanie.

Budowanie ekosystemów firmowych, zgodnie z globalnymi trendami, stopniowo przejdzie do piątego paradygmatu ekonomicznego, w którym analityka skoncentrowana na danych i usługi konsultingowe staną się wyższym priorytetem niż odizolowane, zamknięte rozwiązania o sztwno zdefiniowanych scenariuszach użytkowania.

Era cyfryzacji zmieni układ sił w branży: zamiast polegać na rozwiązaniach dostawców, firmy będą opierać swoją konkurencyjność na zdolności do efektywnego wykorzystywania danych. W rezultacie branża budowlana przejdzie od starszych, sztywnych systemów do elastycznych, adaptacyjnych ekosystemów, w których otwarte standardy i interoperacyjne narzędzia będą podstawą zarządzania projektami. Koniec ery dominacji dostawców aplikacji stworzy nowe środowisko, w którym wartość będzie definiowana nie przez posiadanie zamkniętego kodu źródłowego i wyspecjalizowanych łączników, ale przez zdolność do przekształcania danych w strategiczną przewagę.



ROZDZIAŁ 1.6.

PRZEKSZTAŁCANIE CHAOSU W PORZĄDEK I ZMNIEJSZANIE ZŁOŻONOŚCI

Zbędny kod i zamknięte systemy jako bariera dla poprawy produktywności

W ciągu ostatnich dziesięcioleci zmiany technologiczne w sferze IT były napędzane głównie przez producentów oprogramowania. To oni wyznaczali kierunki rozwoju, określając, które technologie firmy powinny przyjąć, a które pozostawić w tyle. W erze przechodzenia od rozwiązań silosowych do scentralizowanych baz danych i zintegrowanych systemów, dostawcy promowali licencjonowane produkty, zapewniając kontrolę nad dostępem i skalowalnością. Później, wraz z pojawieniem się technologii chmurowych i modeli Software as a Service (SaaS), kontrola ta przekształciła się w model subskrypcji, cementując użytkowników jako lojalnych klientów usług cyfrowych.

Takie podejście doprowadziło do paradoksu: pomimo bezprecedensowej ilości stworzonego kodu programu, tylko niewielka jego część jest faktycznie wykorzystywana. Być może kodu jest setki lub tysiące razy więcej niż potrzeba, ponieważ te same procesy biznesowe są opisywane i powielane w dziesiątkach lub setkach programów na różne sposoby, nawet w ramach tej samej firmy. Jednocześnie koszty rozwoju zostały już opłacone, a koszty te są nie do odzyskania. Mimo to branża nadal powiela ten cykl, tworząc nowe produkty o minimalnej wartości dodanej dla użytkownika końcowego, częściej pod presją oczekiwania rynku niż rzeczywistych potrzeb.

Zgodnie z przewodnikiem szacowania kosztów rozwoju oprogramowania opracowanym przez Defence Acquisition University (DAU) [41], koszt rozwoju oprogramowania może się znacznie różnić w zależności od kilku czynników, w tym złożoności systemu i wybranej technologii. Historycznie, koszty rozwoju w 2008 roku wynosiły około 100 USD za linię kodu źródłowego (SLOC), podczas gdy koszty utrzymania mogą wzrosnąć do 4000 USD za SLOC.

Tylko jeden z komponentów aplikacji CAD - rdzeń geometryczny - może mieć dziesiątki milionów linii kodu (rys. 6.1-5). Podobną sytuację obserwujemy w systemach ERP (rys. 5.4-4), do omówienia złożoności których powrócimy w piątej części książki. Jednak bliższe spojrzenie ujawnia, że większość tego kodu nie dodaje wartości, a jedynie działa jako "listonosz" - mechanicznie przenosząc dane między bazą danych, API, interfejsem użytkownika i innymi tabelami w systemie. Pomimo popularnego mitu o krytycznym znaczeniu tak zwanej logiki biznesowej, trudna rzeczywistość jest znacznie bardziej prozaiczna: nowoczesne bazy kodu są pełne przestarzałych bloków szablonów (starszego kodu), których jedynym celem jest zapewnienie transferu danych między tabelami i komponentami bez wpływu na podejmowanie decyzji lub wydajność biznesową.

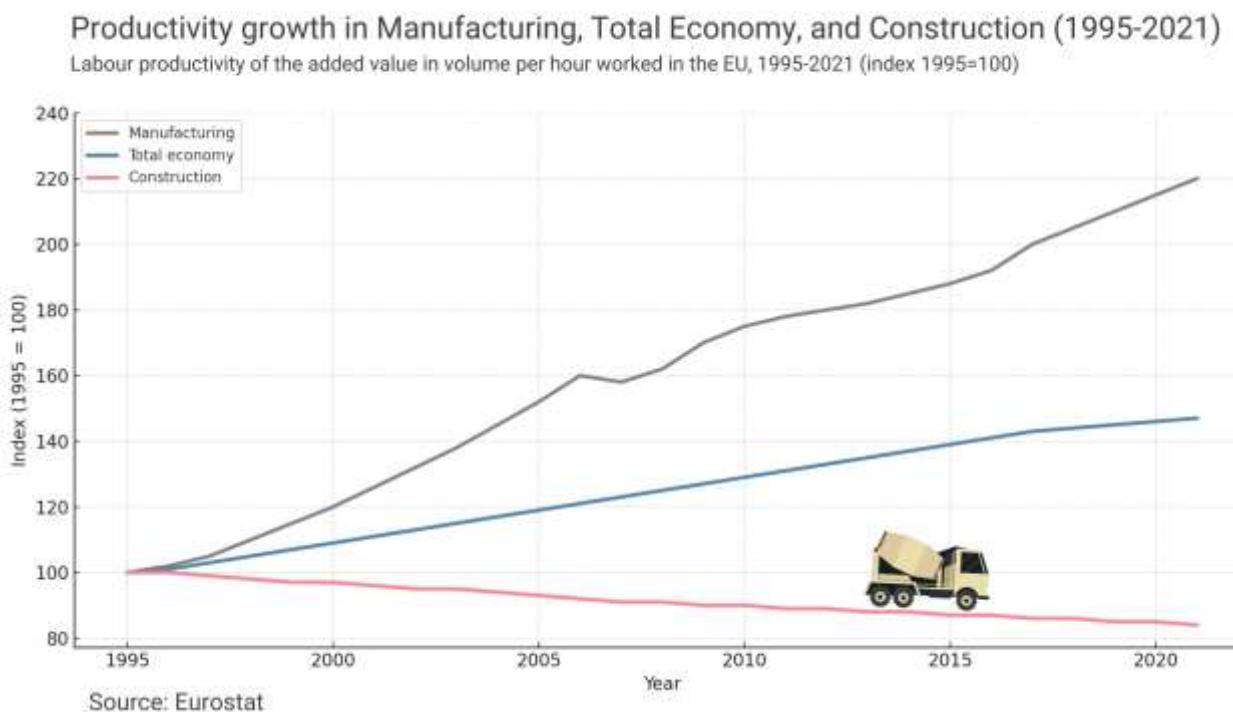
W rezultacie zamknięte rozwiązania przetwarzające dane z różnych źródeł nieuchronnie zamieniają się w zagmatwane "ekosystemy spaghetti". Te złożone, powiązane ze sobą systemy mogą być obsługiwane jedynie przez armię menedżerów pracujących w trybie pół-rutynowych działań. Taka organizacja zarządzania danymi jest nie tylko nieefektywna pod względem zasobów, ale także tworzy krytyczne luki w procesach biznesowych, uzależniając firmę od wąskiego kręgu specjalistów, którzy rozumieją, jak działa ten technologiczny labirynt.

Ciągły wzrost ilości kodu, liczby aplikacji i rosnąca złożoność koncepcji oferowanych przez dostawców doprowadziły do naturalnego rezultatu - wzrostu złożoności ekosystemu IT w budownictwie. Sprawiło to,

że praktyczne wdrażanie cyfryzacji poprzez zwiększanie liczby aplikacji w branży stało się nieefektywne. Oprogramowanie tworzone bez należytej dbałości o potrzeby użytkowników często wymaga znacznych zasobów na wdrożenie i wsparcie, ale nie przynosi oczekiwanych zysków.

Według badania McKinsey "Zwiększenie produktywności w budownictwie" [42] [42], w ciągu ostatnich dwóch dekad globalny wzrost wydajności pracy w budownictwie wynosił średnio zaledwie 1% rocznie, w porównaniu do wzrostu o 2,8% dla całej gospodarki światowej i 3,6% dla przemysłu wytwórczego. W Stanach Zjednoczonych wydajność pracy w budownictwie w przeliczeniu na pracownika spadła o połowę od lat 60. ubiegłego wieku [43].

Rosnąca złożoność systemów, izolacja i zamknięte dane utrudniły komunikację między specjalistami, czyniąc branżę budowlaną jedną z najmniej wydajnych (Rys. 2.2-1). do 22 bilionów dolarów do 2040 roku, co będzie wymagało znacznego wzrostu wydajności.



Rys. 2.2-1 Zamknięte i złożone dane, a w konsekwencji słaba komunikacja między specjalistami doprowadziły branżę budowlaną do jednego z najmniej wydajnych sektorów gospodarki (na podstawie [44], [45]).

Jak podkreślono w badaniu McKinsey (2024) "Zapewnienie wydajności w budownictwie nie jest już opcjonalne", wraz z rosnącym niedoborem zasobów i dążeniem branży do podwojenia tempa wzrostu, budownictwo nie może już sobie pozwolić na utrzymanie obecnego poziomu wydajności [44]. Przewiduje się, że globalne koszty budowy wzrosną z 13 bilionów dolarów w 2023 r. do znacznie wyższych poziomów do końca dekady, co sprawia, że kwestia wydajności jest nie tylko istotna, ale wręcz krytyczna.

Jednym z kluczowych sposobów na poprawę wydajności będzie nieuniknione ujednolicenie i uproszczenie struktur aplikacji i architektur ekosystemów danych. Takie podejście do racjonalizacji wyeliminuje zbędne warstwy abstrakcji i niepotrzebną złożoność, które nagromadziły się przez lata w systemach

korporacyjnych.

Od silosów do pojedynczej hurtowni danych

Im więcej danych gromadzi organizacja, tym trudniej jest wydobyć z nich prawdziwą wartość. Ze względu na fragmentarny charakter przechowywania informacji w odizolowanych silosach, procesy biznesowe współczesnych firm przypominają budowę wieżowca z materiałów przechowywanych w tysiącach różnych magazynów. Nadmiar informacji nie tylko utrudnia dostęp do informacji istotnych z prawnego punktu widzenia, ale także spowalnia podejmowanie decyzji: każdy krok musi być wielokrotnie sprawdzany i potwierdzany.

Każde zadanie lub proces jest na stałe podłączony do oddzielnej tabeli lub bazy danych, a wymiana danych między systemami wymaga złożonych integracji. Błędy i niespójności w jednym systemie mogą powodować awarie łańcucha w innych. Nieprawidłowe wartości, opóźnione aktualizacje i zduplikowane informacje zmuszają pracowników do spędzania znacznej ilości czasu na ręcznym uzgadnianiu i uzgadnianiu danych. W rezultacie organizacja spędza więcej czasu na radzeniu sobie z konsekwencjami fragmentacji niż na opracowywaniu i optymalizacji procesów

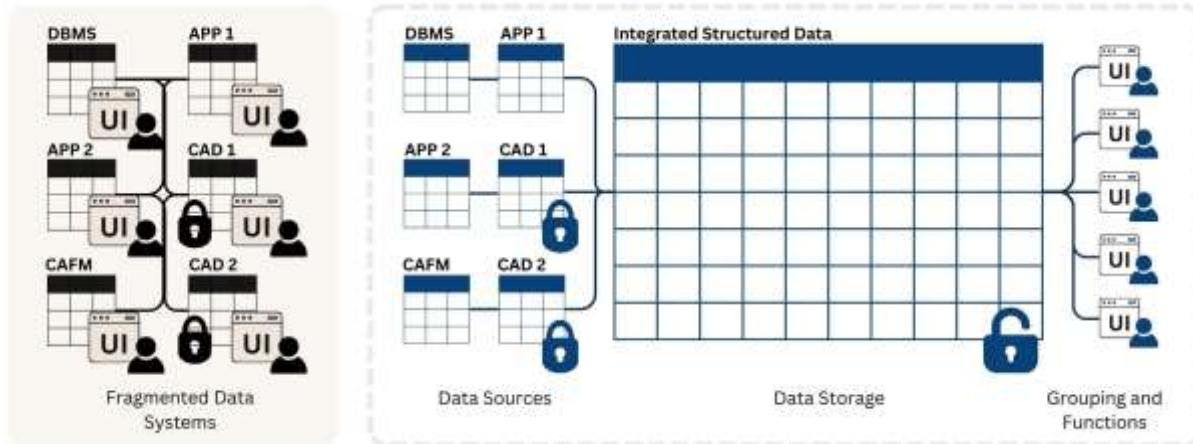
Problem ten jest uniwersalny: niektóre firmy nadal zmagają się z chaosem, podczas gdy inne znajdują rozwiązanie w integracji - przenosząc przepływy informacji do skonsolidowanego systemu przechowywania. Pomyśl o tym jak o jednej dużej tabeli, w której możesz przechowywać wszelkie encje związane z zadaniami, projektami i obiekttami. Zamiast dziesiątek różnych tabel i formatów powstaje jedno spójne repozytorium (Rys. 2.2-2), które umożliwia:

- zminimalizować utratę danych;
- wyeliminować potrzebę ciągłej harmonizacji informacji;
- poprawić dostępność i jakość danych;
- uproszczenie przetwarzania analitycznego i uczenia maszynowego

Doprowadzenie danych do wspólnego standardu oznacza, że niezależnie od źródła, informacje są konwertowane do ujednoliconego formatu nadającego się do odczytu maszynowego. Taka organizacja danych pozwala sprawdzać ich integralność, analizować je w czasie rzeczywistym i szybko wykorzystywać do podejmowania decyzji zarządczych.

Koncepcja zintegrowanych systemów pamięci masowej i ich zastosowania w analityce i uczeniu maszynowym zostanie omówiona bardziej szczegółowo w rozdziale "Big Data Storage and Machine Learning". Tematy modelowania i strukturyzacji danych zostaną szczegółowo omówione w rozdziałach "Przekształcanie danych w ustrukturyzowaną formę" i "Jak standardy zmieniają grę: od losowych plików do

rozbudowanego modelu danych".



Rys. 2.2-2 Integracja danych eliminuje silosy, poprawia dostępność informacji i optymalizuje procesy biznesowe.

Po ustrukturyzowaniu i scaleniu danych kolejnym logicznym krokiem jest ich walidacja. Dzięki pojedynczemu zintegrowanemu repozytorium proces ten jest znacznie uproszczony: koniec z wieloma niespójnymi schematami, powielonymi strukturami i złożonymi relacjami między tabelami. Wszystkie informacje są dostosowane do jednego modelu danych, co eliminuje wewnętrzne niespójności i przyspiesza proces walidacji. Walidacja i zapewnienie jakości danych są podstawowymi aspektami wszystkich procesów biznesowych i przyjrzymy się im bardziej szczegółowo w odpowiednich rozdziałach książki.

Na ostatnim etapie dane są grupowane, filtrowane i analizowane. Stosowane są do nich różne funkcje: agregacja (dodawanie, mnożenie), obliczenia między tabelami, kolumnami lub wierszami (rys. 2.2-4). Praca z danymi staje się sekwencją kroków: gromadzenie, strukturyzacja, walidacja, transformacja, przetwarzanie analityczne i przekazywanie do aplikacji końcowych, w których informacje są wykorzystywane do rozwiązywania praktycznych problemów. Więcej na temat tworzenia takich scenariuszy, automatyzacji kroków i budowania przepływu przetwarzania omówimy w rozdziałach poświęconych procesom ETL i podejściu do potoków danych.

W transformacji cyfrowej nie chodzi więc tylko o uproszczenie obsługi informacji. Chodzi o wyeliminowanie nadmiernej złożoności w zarządzaniu danymi, przejście od chaosu do przewidywalności, od wielu systemów do łatwego w zarządzaniu procesu. Im mniejsza złożoność architektury, tym mniej kodu potrzeba do jej obsługi. A w przeszłości kod jako taki może całkowicie zniknąć, ustępując miejsca intelligentnym agentom, które niezależnie analizują, systematyzują i przekształcą dane.

Zintegrowane systemy pamięci masowej umożliwiają przejście na agentów AI

Im mniej złożone są dane i systemy, tym mniej kodu trzeba napisać i utrzymywać. A najprostszym sposobem na zaoszczędzenie na rozwoju jest całkowite pozbycie się kodu i zastąpienie go danymi. Kiedy rozwój kodu aplikacji przenosi się z kodu na modele danych, nieuchronnie następuje przejście w kierunku podejścia skoncentrowanego na danych (opartego na danych), ponieważ za tymi koncepcjami kryje się zupełnie inny

sposób myślenia.

Kiedy decydujemy się na pracę z danymi w centrum, zaczynamy inaczej postrzegać ich rolę. Dane nie są już tylko "surowcem" dla aplikacji - są teraz fundamentem, wokół którego budowana jest architektura, logika i interakcja.

Tradycyjne podejście do zarządzania danymi zwykle zaczyna się na poziomie aplikacji i w konstrukcji przypomina uciążliwy system biurokratyczny: wielopoziomowe zatwierdzenia, ręczne kontrole, niekończące się wersje dokumentów poprzez odpowiednie oprogramowanie. Wraz z rozwojem technologii cyfrowych coraz więcej firm będzie zmuszonych do przejścia na zasadę minimalizmu - przechowywania i wykorzystywania tylko tego, co jest naprawdę niezbędne i zostanie wykorzystane.

Logika minimalizacji została przejęta przez sprzedawców. Aby uprościć przechowywanie i przetwarzanie danych, praca użytkowników jest przenoszona z aplikacji i narzędzi offline do usług w chmurze i tak zwanych rozwiązań SaaS.

Koncepcja SaaS (Software as a Service, czyli "oprogramowanie jako usługa") jest jednym z kluczowych trendów w nowoczesnej infrastrukturze IT, umożliwiającym użytkownikom dostęp do aplikacji przez Internet bez konieczności instalowania i utrzymywania oprogramowania na własnych komputerach.

Z jednej strony SaaS ułatwił skalowanie, kontrolę wersji i zmniejszył koszty wsparcia i utrzymania, ale z drugiej strony, oprócz zależności od logiki konkretnej aplikacji, całkowicie uzależnił użytkownika od infrastruktury chmury dostawcy. Jeśli usługa ulegnie awarii, dostęp do danych i procesów biznesowych może zostać tymczasowo lub nawet trwale zablokowany. Ponadto wszystkie dane użytkownika podczas pracy z aplikacjami SaaS są przechowywane na serwerach dostawcy, co stwarza zagrożenia dla bezpieczeństwa i zgodności z przepisami. Zmiany w taryfach lub warunkach użytkowania mogą również skutkować wzrostem kosztów lub koniecznością pilnej migracji.

Rozwój agentów AI, LLM i podejścia skoncentrowanego na danych zakwestionował przyszłość aplikacji w ich tradycyjnej formie i realizacji SaaS. Podczas gdy aplikacje i usługi były wcześniej wymagane do zarządzania logiką biznesową i przetwarzania danych, wraz z pojawieniem się agentów AI funkcje te mogą zostać przeniesione do inteligentnych systemów, które pracują bezpośrednio z danymi.

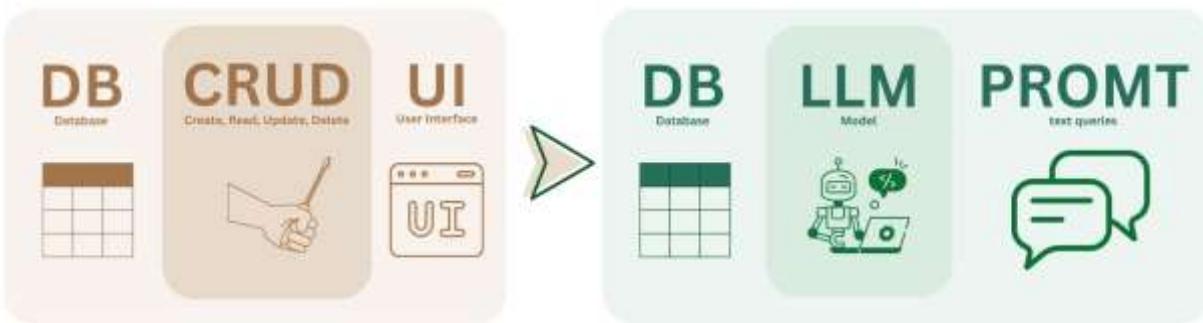
Dlatego też w działach IT i na szczeblu kierowniczym coraz częściej dyskutuje się o architekturach hybrydowych, w których agenci AI i rozwiązania lokalne uzupełniają usługi w chmurze, zmniejszając zależność od platform SaaS.

Podejście, które przyjmujemy, uznaje, że tradycyjne aplikacje biznesowe lub aplikacje SaaS mogą ulec radykalnej zmianie w erze agentów. Aplikacje te są zasadniczo bazami danych CRUD [utwórz, odczytaj, zaktualizuj i usuń] z logiką biznesową. Jednak w przyszłości logika ta zostanie przejęta przez agentów AI [46].

- Satya Nadella, dyrektor generalny Microsoft, 2024 r.

Podejście skoncentrowane na danych i wykorzystanie agentów AI/LLM może ograniczyć zbędne procesy, a tym samym zmniejszyć obciążenie pracowników. Gdy dane są odpowiednio zorganizowane, łatwiej jest je analizować, wizualizować i wykorzystywać do podejmowania decyzji. Zamiast niekończących się raportów i kontroli, specjaliści uzyskują dostęp do aktualnych informacji za pomocą kilku kliknięć lub z pomocą agentów LLM automatycznie w postaci gotowych dokumentów i dashboardów.

W manipulowaniu danymi pomogą nam narzędzia sztucznej inteligencji (AI) oraz czaty LLM. W ostatnich latach nastąpił trend odchodzenia od tradycyjnych operacji CRUD (create, read, update, delete) w kierunku wykorzystania dużych modeli językowych (LLM) do zarządzania danymi. Modele LLM są w stanie interpretować język naturalny i automatycznie generować odpowiednie zapytania do baz danych, co upraszcza interakcję z systemami zarządzania danymi (rys. 2.2-3).



Rys. 2.2-3 AI zastąpi i zintegruje rozwiązania pamięci masowej i baz danych, stopniowo wypierając tradycyjne aplikacje i operacje CRUD.

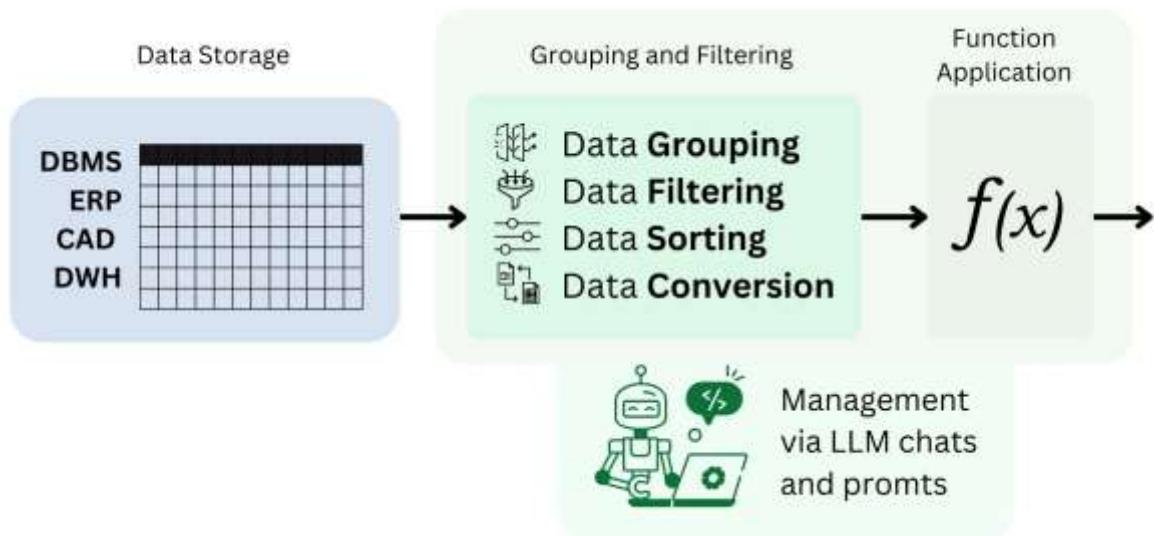
W ciągu najbliższych 3-6 miesięcy AI będzie pisać 90% kodu, a w ciągu 12 miesięcy prawie cały kod może być generowany przez AI [47].

- Dario Amodei, dyrektor generalny LLM Anthropic, marzec 2025 r.

Pomimo szybkiego rozwoju narzędzi do rozwoju AI (np. GitHub Copilot), w 2025 roku deweloperzy nadal będą odgrywać kluczową rolę w tym procesie. Agenci AI stają się coraz bardziej użytecznymi asystentami: automatycznie interpretują zapytania użytkownika, generują zapytania SQL i Pandas (więcej na ten temat w kolejnych rozdziałach) lub piszą kod do analizy danych. W ten sposób sztuczna inteligencja stopniowo zastępuje tradycyjne interfejsy użytkownika aplikacji.

Rozprzestrzenianie się modeli sztucznej inteligencji, takich jak modele językowe, będzie napędzać rozwój architektur hybrydowych. Zamiast całkowicie rezygnować z rozwiązań chmurowych i produktów SaaS, możemy zaobserwować integrację usług chmurowych z lokalnymi systemami zarządzania danymi. Na przykład uczenie federacyjne umożliwia tworzenie potężnych modeli AI bez konieczności przenoszenia wrażliwych danych do chmury. W ten sposób firmy mogą zachować kontrolę nad swoimi danymi,

jednocześnie uzyskując dostęp do zaawansowanych technologii.



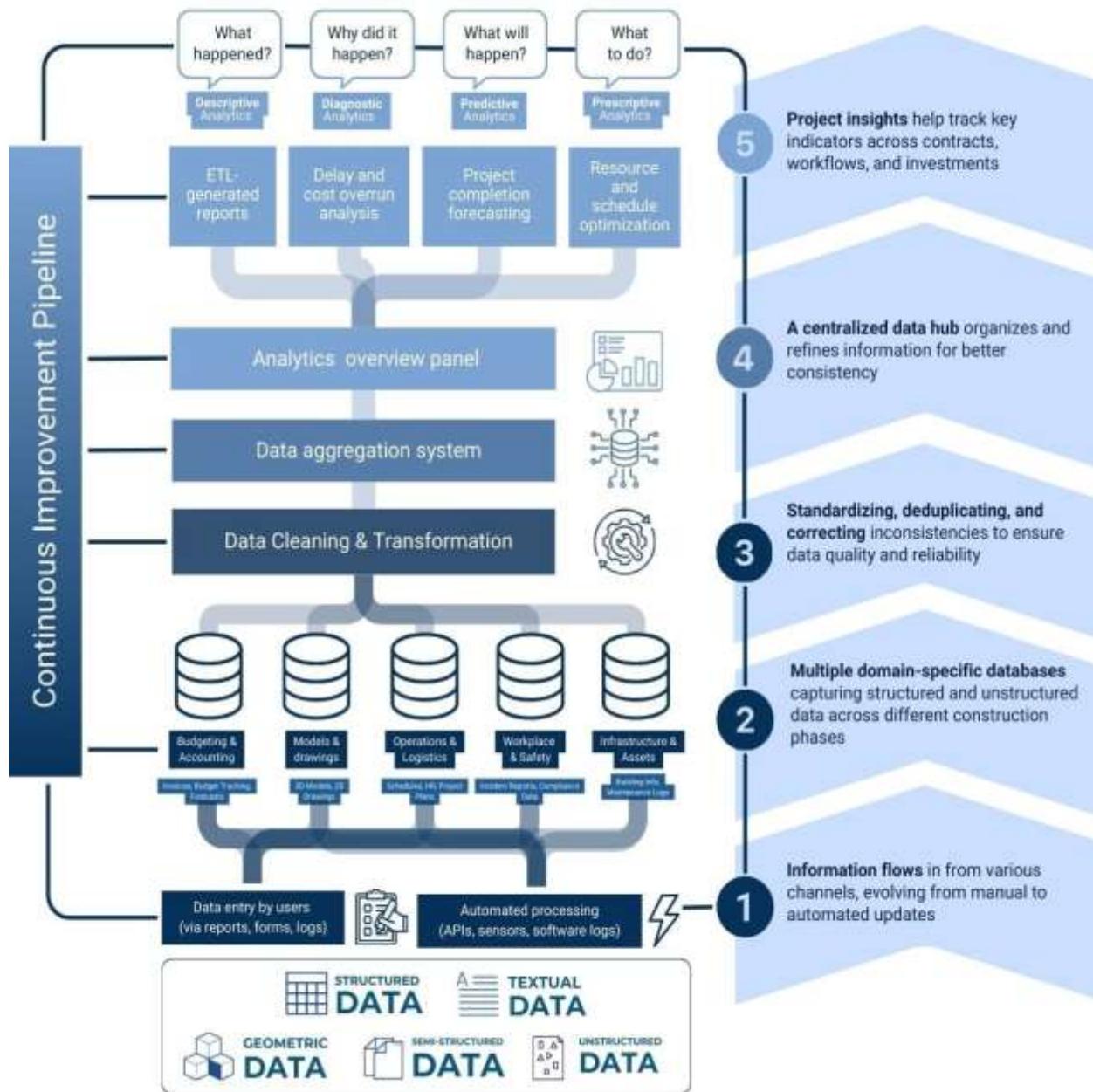
Rys. 2.2-4 Podstawowe operacje grupowania, filtrowania i sortowania, po których następuje zastosowanie funkcji, będą obsługiwane przez czaty LLM.

Przyszłość branży budowlanej będzie opierać się na połączeniu rozwiązań lokalnych, mocy chmury i inteligentnych modeli współpracujących ze sobą w celu stworzenia wydajnych i bezpiecznych systemów zarządzania danymi. LLM umożliwi użytkownikom bez głębokiej wiedzy technicznej interakcję z bazami danych i hurtowniami danych poprzez formułowanie zapytań w języku naturalnym. Więcej informacji na temat agentów LLM i AI oraz ich działania znajduje się w rozdziale "Agenci LLM i ustrukturyzowane formaty danych".

Właściwie zorganizowane dane i proste, łatwe w użyciu narzędzia analityczne LLM nie tylko ułatwiają pracę z informacjami, ale także pomogą zminimalizować błędy, zwiększyć wydajność i zautomatyzować procesy.

Od gromadzenia danych do podejmowania decyzji: droga do automatyzacji

W późniejszych częściach książki przyjrzymy się szczegółowo, w jaki sposób specjalisi współdziałają ze sobą i jak dane stają się podstawą podejmowania decyzji, automatyzacji i wydajności operacyjnej. Rys. 2.2-5 przedstawia przykładowy diagram pokazujący sekwencję etapów przetwarzania danych w podejściu skoncentrowanym na danych. Schemat ten ilustruje potok ciągłego doskonalenia), którego części zostaną szczegółowo omówione w dalszej części książki.



Rys. 2.2-5 Przykład potoku ciągłego doskonalenia danych: przepływ przetwarzania i analizy danych w projektach budowlanych.

System opisujący procesy biznesowe średniej wielkości firmy zbudowany jest na zasadzie wielopoziomowej. Obejmuje on: gromadzenie danych, czyszczenie, agregację, przetwarzanie analityczne i podejmowanie decyzji na podstawie wyników. Wszystkie te etapy przeanalizujemy w dalszej części książki - zarówno w kontekście teoretycznym, jak i na praktycznych przykładach:

- Na pierwszym poziomie odbywa się **wprowadzanie danych** (rys. 3.1-1). Informacje są odbierane zarówno ręcznie (poprzez raporty, formularze, dzienniki), jak i w formie zautomatyzowanej (z API, czujników, systemów oprogramowania). Dane mogą mieć różną strukturę: geometryczną, tekstową, nieustrukturyzowaną. Na tym etapie istnieje potrzeba standaryzacji, strukturyzacji i

ujednolicenia przepływów informacji.

- Kolejnym poziomem jest **przetwarzanie i przekształcanie danych**. Obejmuje on procesy czyszczenia, usuwania duplikatów, korygowania błędów i przygotowywania informacji do dalszej analizy (Rys. 4.2-5). Ten etap jest krytyczny, ponieważ jakość analityki zależy bezpośrednio od czystości i dokładności danych.
- **Dane trafiają** następnie do **wyspecjalizowanych tabel, ramek danych lub baz danych** podzielonych według obszarów funkcjonalnych: budżetowanie i księgowość, modele i rysunki, logistyka, bezpieczeństwo i infrastruktura. Taki podział pozwala na łatwy dostęp i analizę krzyżową informacji.
- Dane są następnie **agregowane i wyświetlane na pulpicie analitycznym** (prezentacja). Stosowane są tu metody analityki opisowej, diagnostycznej, predykcyjnej i preskryptywnej. Pomaga to odpowiedzieć na kluczowe pytania (rys. 1.1-4): co się stało, dlaczego tak się stało, co stanie się w przyszłości i jakie działania należy podjąć. Na przykład system może identyfikować opóźnienia, przewidywać zakończenie projektów lub optymalizować zasoby.
- Wreszcie, ostatni poziom generuje **wnioski analityczne i kluczowe wskaźniki**, które pomagają monitorować realizację kontraktów, zarządzać inwestycjami i usprawniać procesy biznesowe (rys. 7.4-2). Informacje te stają się podstawą do podejmowania decyzji i strategii rozwoju firmy.

Podobnie, dane przechodzą od etapu gromadzenia do wykorzystania w zarządzaniu strategicznym. W kolejnych częściach książki przyjrzymy się szczegółowo każdemu etapowi, koncentrując się na typach danych, technikach przetwarzania danych, narzędziach analitycznych oraz rzeczywistych przypadkach wykorzystania tych podejść w branży budowlanej.

Kolejne kroki: przekształcenie chaosu w zarządzalny system

W tej części zbadaliśmy wyzwania związane z silosami informacyjnymi i przyjrzaliśmy się wpływom nadmiernej złożoności systemu na wydajność biznesową, analizując przejście od czwartej rewolucji przemysłowej do piątej, w której dane, a nie aplikacje, mają kluczowe znaczenie. Zobaczyliśmy, jak silosowe systemy informatyczne tworzą bariery dla dzielenia się wiedzą, a ciągła złożoność środowiska IT zmniejsza produktywność i hamuje innowacje w branży budowlanej.

Podsumowując tę część, warto podkreślić główne praktyczne kroki, które pomogą ci zastosować omówione podejścia w codziennych zadaniach:

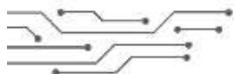
- **Wizualizacja krajobrazu informacyjnego**
 - Stwórz wizualną mapę źródeł danych (Miro, Figma, Canva), z którymi regularnie pracujesz.
 - Dodaj do tej mapy systemy i aplikacje, z których korzystasz w swojej pracy.
 - Identyfikacja potencjalnie zduplikowanych funkcji i zbędnych rozwiązań
 - Identyfikacja krytycznych punktów, w których może dojść do utraty lub uszkodzenia danych podczas transmisji między systemami.
- **Wdrożenie spersonalizowanych praktyk zarządzania danymi**
 - Przeniesienie punktu ciężkości z aplikacji na dane jako kluczowy zasób w procesach
 - Dokumentowanie źródeł danych i metodologii przetwarzania w celu zapewnienia

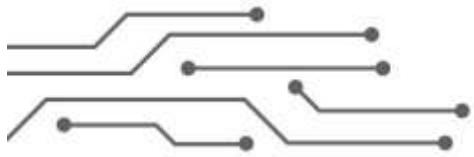
przejrzystości

- Opracowanie mechanizmów oceny i poprawy jakości danych
- Należy dążyć do tego, aby dane były wprowadzane raz i wykorzystywane wielokrotnie - jest to podstawa efektywnej organizacji procesów.
- Promowanie w zespole podejścia skoncentrowanego na danych (opartego na danych).
 - Sugerowanie stosowania standardowych i jednolitych formatów wymiany danych peer-to-peer.
 - Regularne podnoszenie kwestii związanych z jakością i dostępnością danych na spotkaniach zespołu.
 - Poznaj alternatywy Open Source dla narzędzi, których używasz do rozwiązywania swoich problemów.

Zacznij od małego - wybierz jeden konkretny proces lub zestaw danych, który ma kluczowe znaczenie dla Twojej pracy i zastosuj do niego podejście skoncentrowane na danych, przenosząc nacisk z narzędzi na dane. Osiągając sukces w pojedynczym pilotażu, nie tylko zdobędziesz praktyczne doświadczenie, ale także wyraźnie zademonstrujesz zespołowi korzyści płynące z nowej metodologii. Jeśli w trakcie realizacji większości z tych kroków pojawią się pytania, można zwrócić się o wyjaśnienie i pomoc do dowolnego aktualnego LLM.

W kolejnych częściach książki przejdziemy do bardziej szczegółowego spojrzenia na techniki strukturyzacji i harmonizacji danych oraz zbadamy praktyczne podejścia do integracji heterogenicznych informacji. Szczególna uwaga zostanie poświęcona przejściu od rozproszonych silosów do ujednolicionych ekosystemów danych, które odgrywają kluczową rolę w cyfrowej transformacji branży budowlanej.





II CZĘŚĆ

STRUKTURA DANYCH W PROCESACH BIZNESOWYCH W BRANŻY BUDOWLANEJ

W trzeciej części przedstawiono kompleksowe zrozumienie typologii danych w budownictwie i metod ich efektywnej organizacji. Przeanalizowano charakterystykę i specyfikę pracy z danymi ustrukturyzowanymi, nieustrukturyzowanymi, częściowo ustrukturyzowanymi, tekstowymi i geometrycznymi w kontekście projektów budowlanych. Dokonano przeglądu nowoczesnych formatów przechowywania i protokołów wymiany informacji między różnymi systemami stosowanymi w branży. Opisano praktyczne narzędzia i techniki przekształcania danych wieloformatowych w jedno ustrukturyzowane środowisko, w tym sposoby integracji danych CAD (BIM). Zaproponowano podejścia do zapewnienia jakości danych poprzez standaryzację i walidację, co ma kluczowe znaczenie dla dokładności obliczeń konstrukcyjnych. Szczegółowo przeanalizowano praktyczne aspekty wykorzystania nowoczesnych technologii (Python Pandas, LLM -models) z przykładami kodu do rozwiązywania typowych problemów w branży budowlanej. Uzasadniono wartość utworzenia centrum kompetencji (CoE) jako struktury organizacyjnej do koordynacji i standaryzacji podejść do zarządzania informacją.

ROZDZIAŁ 2.1.

TYPY DANYCH W BUDOWNICTWIE

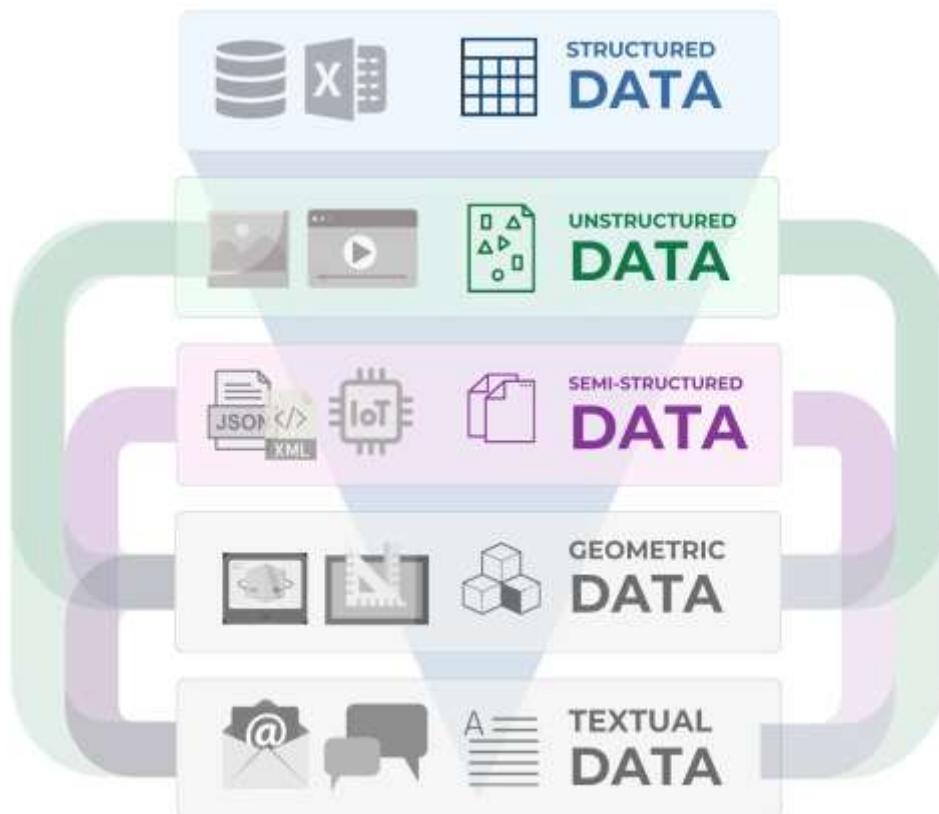
Najważniejsze typy danych w branży budowlanej

W nowoczesnej branży budowlanej systemy, aplikacje i hurtownie danych firm są aktywnie wypełniane informacjami i danymi różnych typów i formatów (rys. 3.1-1). Przyjrzyjmy się bliżej głównym typom danych, które tworzą krajobraz informacyjny nowoczesnej firmy działającej w branży budowlanej:

- **Dane ustrukturyzowane:** dane te mają wyraźną strukturę organizacyjną, np. arkusze kalkulacyjne Excel i relacyjne bazy danych.
- **Dane nieustrukturyzowane:** są to informacje, które nie są zorganizowane zgodnie ze ścisłymi zasadami. Przykładami takich danych są tekst, wideo, zdjęcia i nagrania audio.
- **Dane luźno ustrukturyzowane:** dane te zajmują pozycję pośrednią między danymi ustrukturyzowanymi i nieustrukturyzowanymi. Zawierają elementy struktury, ale struktura ta nie zawsze jest jasna lub często opisana za pomocą różnych schematów. Przykładami danych częściowo ustrukturyzowanych w budownictwie są: specyfikacje techniczne, dokumentacja projektowa lub raporty z postępu prac.
- **Dane tekstowe:** obejmują wszystko, co pochodzi z komunikacji ustnej i pisemnej, takie jak e-maile, transkrypcje spotkań i spotkań.
- **Dane geometryczne:** dane te pochodzą z programów CAD, w których specjalisci tworzą dane geometryczne elementów projektu w celu wizualizacji, potwierdzenia wartości objętości lub sprawdzenia kolizji.

Należy zauważyć, że dane geometryczne i tekstowe (alfanumeryczne) nie stanowią odrębnej kategorii, ale mogą występować we wszystkich trzech typach danych. Na przykład dane geometryczne mogą być częścią zarówno danych ustrukturyzowanych (parametryczne formaty CAD), jak i nieustrukturyzowanych (zeskanowane rysunki). Dane tekstowe mogą podobnie być zarówno zorganizowane w bazach danych (dane ustrukturyzowane), jak i istnieć jako dokumenty bez wyraźnej struktury.

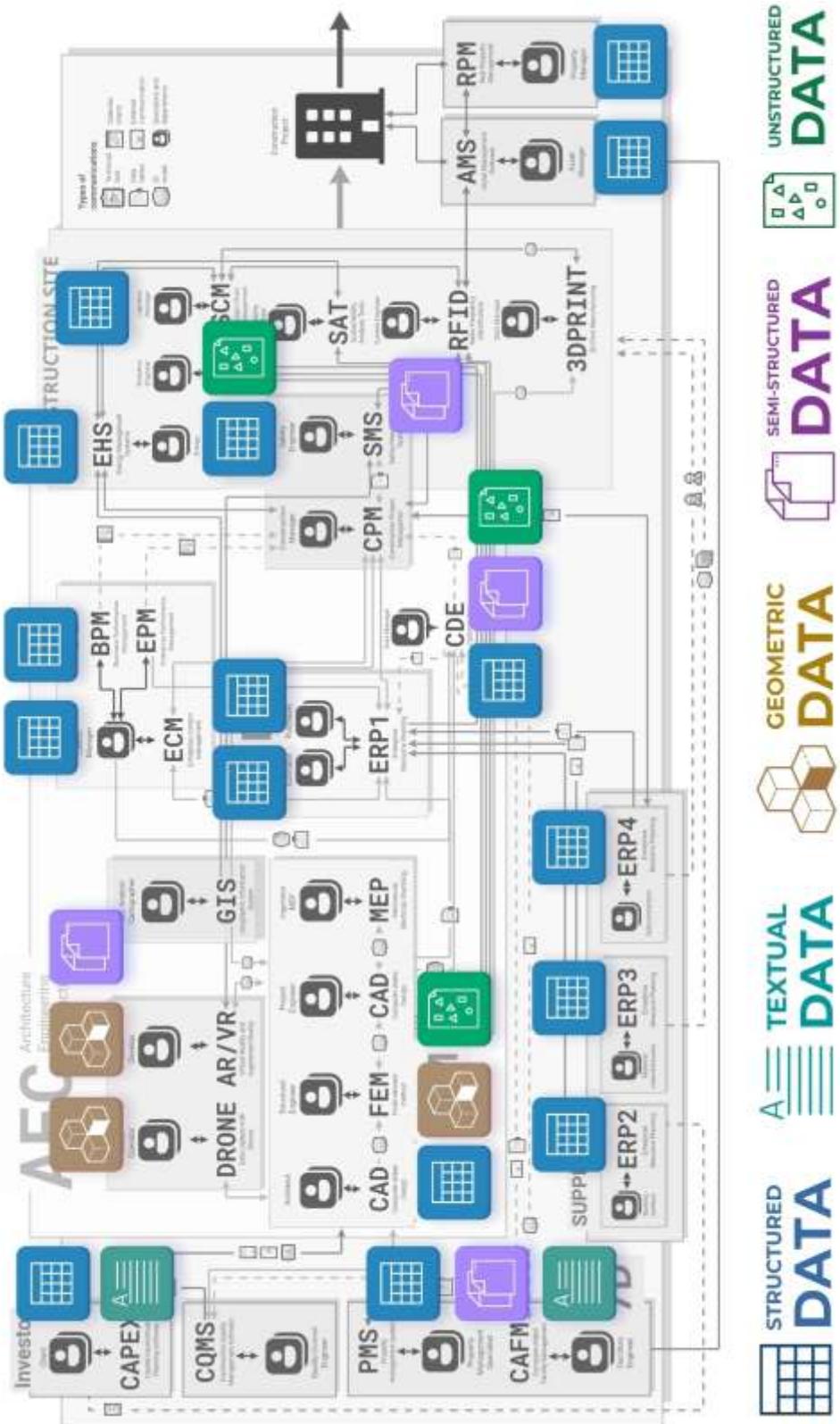
Każdy rodzaj danych w firmie budowlanej jest unikalnym elementem mozaiki zasobów informacyjnych firmy. Od nieustrukturyzowanych danych, takich jak zdjęcia z placów budowy i nagrania audio ze spotkań, po ustrukturyzowane rekordy, w tym tabele i bazy danych, każdy element odgrywa ważną rolę w kształtowaniu krajobrazu informacyjnego firmy.



Rys. 3.1-1 Inżynierowie i menedżerowie danych muszą nauczyć się pracować ze wszystkimi rodzajami danych wykorzystywanych w branży budowlanej.

Oto przykładowa lista tylko niektórych systemów i powiązanych typów danych (Rys. 3.1-2) wykorzystywanych w budownictwie:

- **ERP** (Enterprise Resource Planning) - obsługuje ogólnie ustrukturyzowane dane, aby pomóc w zarządzaniu zasobami przedsiębiorstwa i integracji różnych procesów biznesowych.
- **CAD** (Computer-Aided Design) w połączeniu z **BIM** (Building Information Modeling) - wykorzystuje geometryczne i częściowo ustrukturyzowane dane do projektowania i modelowania projektów budowlanych, zapewniając dokładność i spójność informacji w fazie projektowania.
- **GIS** (Systemy Informacji Geograficznej) - pracuje z danymi geometrycznymi i strukturalnymi w celu tworzenia i analizowania danych kartograficznych i relacji przestrzennych.
- **RFID** (Radio-Frequency Identification) - wykorzystuje częściowo ustrukturyzowane dane do skutecznego śledzenia materiałów i sprzętu na placu budowy za pomocą identyfikacji radiowej.
- **ECM** (Engineering Content Management) to system zarządzania danymi i dokumentacją inżynierską, w tym danymi częściowo ustrukturyzowanymi i nieustrukturyzowanymi, takimi jak rysunki techniczne i dokumenty projektowe.

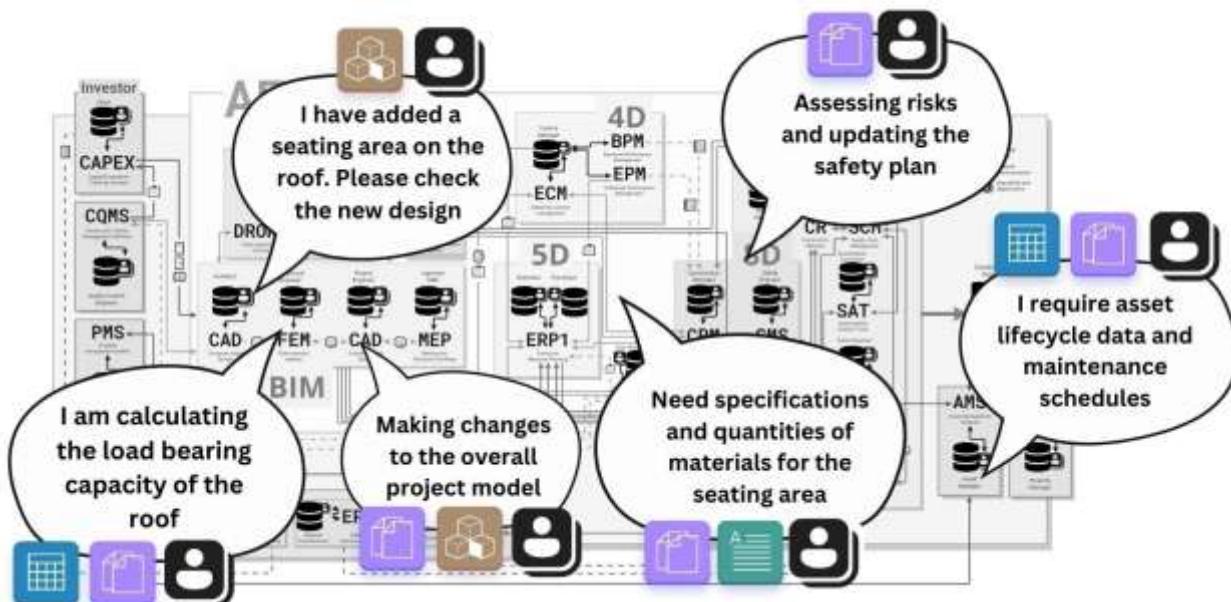


Rys. 3.1-2 Różne formaty i dane wypełniają różne systemy, wymagając przetłumaczenia na formę odpowiednią do złożonej integracji.

Te i wiele innych systemów firmy zarządza szerokim zakresem danych, od ustrukturyzowanych danych tabelarycznych po złożone modele geometryczne, zapewniając zintegrowaną współpracę w procesach projektowania, planowania i zarządzania budową.

W przykładzie uproszczonego dialogu (rys. 3.1-3) różne rodzaje danych są wymieniane między specjalistami ds. projektów budowlanych:

- ⦿ **Architekt:** "Biorąc pod uwagę życzenia klienta, dodałem miejsce do siedzenia na dachu. Zapraszam do zapoznania się z nowym projektem" (dane geometryczne - model).
- ⦿ **Inżynier budowlany:** "Projekt został odebrany. Obliczam nośność dachu dla nowej strefy rekreacyjnej" (dane ustrukturyzowane i częściowo ustrukturyzowane - tabele obliczeniowe).
- ⦿ **Kierownik ds. zakupów:** "Potrzebujesz specyfikacji i ilości materiałów na teren rekreacyjny, aby zorganizować zakup" (dane tekstowe i częściowo ustrukturyzowane - listy i specyfikacje).
- ⦿ **Inżynier ds. zdrowia i bezpieczeństwa:** "Otrzymałem dane dotyczące nowego obszaru. Oceniam ryzyko i aktualizuję plan bezpieczeństwa" (dane częściowo ustrukturyzowane - dokumenty i plany).
- ⦿ **Specjalista ds. modelowania BIM:** "Wprowadzanie zmian w ogólnym modelu projektu w celu dostosowania dokumentacji roboczej" (dane geometryczne i dane półstrukturalne).
- ⦿ **Kierownik projektu:** "Włączam nową strefę odpoczynku do harmonogramu prac. Aktualizuję harmonogramy i zasoby w systemie zarządzania projektami" (dane ustrukturyzowane i częściowo ustrukturyzowane - harmonogramy i plany).
- ⦿ **Specjalista ds. utrzymania obiektów (FM):** "Przygotowuję dane do przyszłego utrzymania terenów rekreacyjnych i wprowadzam je do systemu zarządzania nieruchomościami" (dane ustrukturyzowane i częściowo ustrukturyzowane - instrukcje i plany utrzymania).



Rys. 3.1-3 Komunikacja między specjalistami odbywa się zarówno na poziomie tekstu, jak i danych.

Każdy profesjonalista pracuje z różnymi typami danych, aby zapewnić efektywną współpracę zespołu i sukces projektu. Zrozumienie różnic między danymi ustrukturyzowanymi, częściowo ustrukturyzowanymi i nieustrukturyzowanymi pozwala rozpoznać wyjątkową rolę, jaką każdy typ odgrywa w cyfrowych procesach biznesowych. Ważne jest nie tylko, aby wiedzieć, że istnieją różne formy danych, ale także zrozumieć, w jaki

sposób, gdzie i dlaczego są one wykorzystywane.

Jeszcze niedawno pomysł łączenia tak różnorodnych danych wydawał się ambitny, ale trudny do zrealizowania. Dziś jest to już część codziennej praktyki. Integracja danych o różnych schematach i strukturach stała się integralną częścią nowoczesnej architektury systemów informatycznych.

W kolejnych rozdziałach przyjrzymy się szczegółowo kluczowym standardom i podejściom, które umożliwiają łączenie ustrukturyzowanych, częściowo ustrukturyzowanych i nieustrukturyzowanych danych w jeden spójny widok. Szczególna uwaga zostanie poświęcona ustrukturyzowanym danym i relacyjnym bazom danych jako głównym mechanizmom przechowywania, przetwarzania i analizowania informacji w branży budowlanej.

Dane strukturalne

W branży budowlanej informacje pochodzą z wielu źródeł - rysunków, specyfikacji, harmonogramów i raportów. Aby skutecznie zarządzać tym przepływem informacji, należy je ustrukturyzować. Dane strukturalne umożliwiają organizowanie informacji w wygodnej, czytelnej i dostępnej formie.

Według piątego rocznego raportu JB Knowledge na temat technologii budowlanych [17], 67% specjalistów ds. zarządzania projektami budowlanymi śledzi i ocenia wydajność pracy ręcznej lub za pomocą arkuszy kalkulacyjnych.

Niektóre z najpopularniejszych formatów danych strukturalnych to XLSX i CSV. Są one szeroko stosowane do przechowywania, przetwarzania i analizowania informacji w arkuszach kalkulacyjnych. W takich arkuszach kalkulacyjnych dane są prezentowane w postaci wierszy i kolumn, co ułatwia ich odczyt, edycję i analizę.

XLSX, format stworzony przez Microsoft, opiera się na wykorzystaniu struktur XML i jest archiwizowany przy użyciu algorytmu ZIP. Główne cechy formatu:

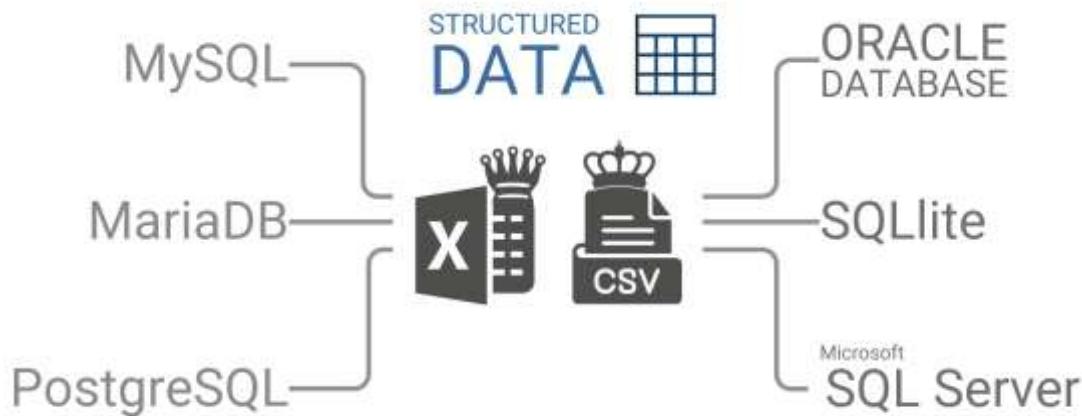
- Obsługa złożonych formuł, wykresów i makr.
- Możliwość przechowywania danych w różnych arkuszach, a także formatowania informacji.
- Zoptymalizowany dla Microsoft Excel, ale kompatybilny z innymi pakietami biurowymi.

Format CSV to zwykły plik tekstowy, w którym wartości są oddzielone przecinkami, średnikami lub innymi znakami ograniczającymi. Główne zalety:

- Uniwersalna kompatybilność z różnymi programami i systemami operacyjnymi.
- Łatwy import/eksport do baz danych i systemów analitycznych.
- Łatwe przetwarzanie nawet w edytorech tekstu.

Jednak CSV nie obsługuje formuł i formatowania, więc jego głównym zastosowaniem jest wymiana danych między systemami i masowe aktualizacje informacji. Ze względu na swoją wszechstronność i niezależność od platformy, CSV stał się popularnym narzędziem do przesyłania danych w heterogenicznych środowiskach IT.

Dwa formaty XLSX i CSV działają jako łącznik między różnymi systemami zajmującymi się danymi strukturalnymi (Rys. 3.1-4). Są one szczególnie przydatne w zadaniach, w których ważna jest czytelność, ręczna edycja i podstawowa kompatybilność



Rys. 3.1-4 Formaty XLSX i CSV są łącznikiem między różnymi systemami, które pracują z danymi strukturalnymi.

Niezależność od platformy sprawia, że CSV jest najpopularniejszym formatem przesyłania danych w heterogenicznych środowiskach i systemach IT.

Jednak formaty XLSX i CSV nie są przeznaczone do wysokowydajnych obliczeń lub długoterminowego przechowywania dużych ilości danych. Do takich celów wykorzystywane są bardziej nowoczesne formaty strukturalne, takie jak Apache Parquet, Apache ORC, Feather, HDF5. Formaty te zostaną omówione bardziej szczegółowo w rozdziale "Przechowywanie dużych ilości danych: analiza popularnych formatów i ich skuteczności" w części 9 tej książki

W praktyce Excel z formatem XLSX jest częściej używany do małych zadań i automatyzacji rutynowych procesów. Bardziej złożone scenariusze wymagają użycia systemów zarządzania danymi, takich jak ERP, PMIS CAFM, CPM, SCM i innych (rys. 3.2-1). Systemy te przechowują ustrukturyzowane dane, na których opiera się organizacja i zarządzanie przepływem informacji w firmie.

Nowoczesne systemy zarządzania danymi wykorzystywane w branży budowlanej opierają się na ustrukturyzowanych danych zorganizowanych w formie tabel. W celu zapewnienia niezawodnego, skalowalnego i całościowego zarządzania dużymi ilościami informacji, twórcy aplikacji i systemów zwracają się ku systemom zarządzania relacyjnymi bazami danych (RDBMS).

Relacyjne bazy danych RDBMS i język zapytań SQL

Aby efektywnie przechowywać, przetwarzać i analizować dane, **relacyjne bazy danych (RDBMS)** są

systemami przechowywania danych, które organizują informacje w tabele ze zdefiniowanymi relacjami między nimi.

Dane zorganizowane w bazach danych (RDBMS) to nie tylko informacje cyfrowe; są one podstawą transakcji i interakcji między różnymi systemami.

Oto kilka najpopularniejszych systemów zarządzania relacyjnymi bazami danych (RDBMS) (Rys. 3.1-5):

- **MySQL** (Open Source) jest jednym z najpopularniejszych RDBMS, który jest częścią stosu LAMP (Linux, Apache, MySQL, PHP /Perl/Python). Jest szeroko stosowany w tworzeniu stron internetowych ze względu na swoją prostotę i wysoką wydajność.
- **PostgreSQL** (Open Source) to potężny system obiektowo-relacyjny znany ze swojej niezawodności i zaawansowanych funkcji. Nadaje się do złożonych rozwiązań korporacyjnych.
- **Microsoft SQL Server** to komercyjny system firmy Microsoft, który jest szeroko stosowany w środowiskach korporacyjnych ze względu na jego integrację z innymi produktami firmy i wysoki poziom bezpieczeństwa.
- **Oracle Database** to jeden z najpotężniejszych i najbardziej niezawodnych systemów DBMS wykorzystywanych w dużych przedsiębiorstwach i aplikacjach o znaczeniu krytycznym.
- **IBM DB2** - przeznaczony dla dużych korporacji, zapewniający wysoką wydajność i odporność na awarie.
- **SQLite** (Open Source) to lekka wbudowana baza danych, idealna dla aplikacji mobilnych i samodzielnych systemów, takich jak oprogramowanie do projektowania CAD (BIM).

Popularne systemy zarządzania bazami danych w branży budowlanej - MySQL, PostgreSQL, Microsoft SQL Server, Oracle® Database, IBM® DB2 i SQLite - pracują z danymi strukturalnymi. Wszystkie te systemy DBMS są potężnymi i elastycznymi rozwiązaniami do zarządzania szeroką gamą procesów biznesowych i aplikacji, od małych stron internetowych po duże systemy korporacyjne (rys. 3.2-1).

Według Statista [48], relacyjne systemy zarządzania bazami danych (RDBMS) stanowią około 72% wszystkich DBMS w użyciu w 2022 roku.

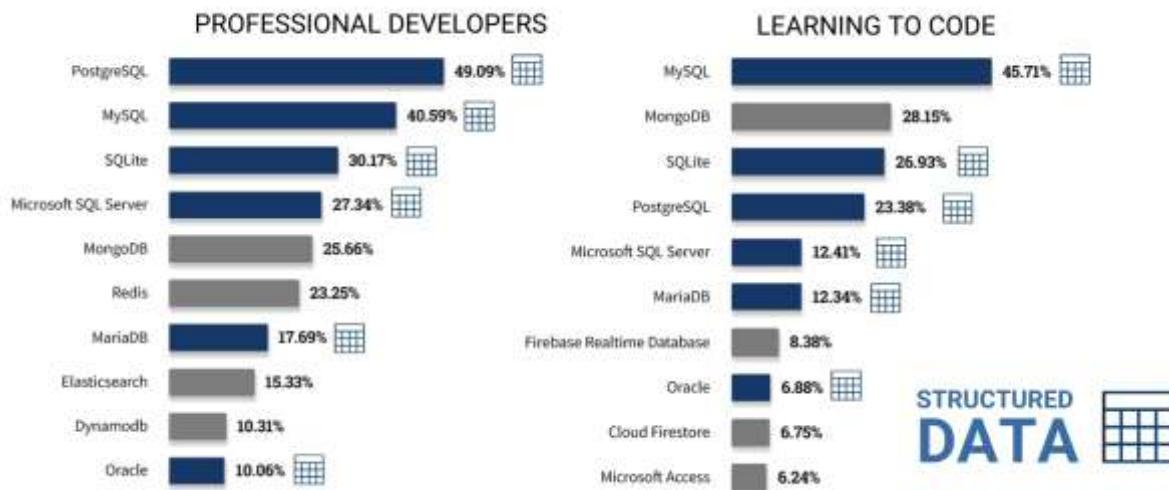
	Rank			DBMS	Database Model	Open Source vs Commercial
	Mar2025	Feb2025	Mar2024			
	1.	1.	1.	Oracle®	Relational, Multi-model	Commercial
	2.	2.	2.	MySQL	Relational, Multi-model	Open Source
	3.	3.	3.	Microsoft® SQL Server	Relational, Multi-model	Commercial
	4.	4.	4.	PostgreSQL	Relational, Multi-model	Open Source
	5.	5.	5.	MongoDB	Document, Multi-model	Open Source
	6.	7.	9.	Snowflake®	Relational	Commercial
	7.	6.	6.	Redis®	Key-value, Multi-model	Open Source
	8.	8.	7.	Elasticsearch®	Multi-model	Open Source
	9.	9.	8.	IBM Db2	Relational, Multi-model	Commercial
	10.	10.	10.	SQLite	Relational	Open Source
	11.	11.	12.	Apache Cassandra®	Multi-model	Open Source
	12.	12.	11.	Microsoft Access®	Relational	Open Source
	13.	13.	17.	Databricks®	Multi-model	Commercial
	14.	14.	13.	MariaDB	Relational, Multi-model	Open Source
	15.	15.	14.	Splunk	Search engine	Commercial
	16.	16.	16.	Amazon DynamoDB	Multi-model	Commercial
	17.	17.	15.	Microsoft Azure SQL	Relational, Multi-model	Commercial

Rys. 3.1-5 Popularność korzystania ze strukturalnych baz danych (zaznaczonych na niebiesko) w rankingu SZBD (na podstawie [49]).

Instalacja baz danych open source jest dość łatwa - nawet bez rozległej wiedzy technicznej. Systemy open source, takie jak PostgreSQL, MySQL czy SQLite, są dostępne za darmo i działają na większości systemów operacyjnych: Windows, macOS i Linux. Wystarczy wejść na oficjalną stronę projektu, pobrać instalator i postępować zgodnie z instrukcjami. W większości przypadków instalacja zajmuje nie więcej niż 10-15 minut. Jedną z takich baz danych zamodelujemy i utworzymy w czwartej części książki (rys. 4.3-8).

Jeśli Twoja firma korzysta z usług w chmurze (na przykład Amazon Web Services, Google Cloud lub Microsoft Azure), możesz wdrożyć bazę danych za pomocą kilku kliknięć - platforma zaoferuje Ci gotowe szablony do instalacji. Dzięki otwartości kodu, takie bazy danych można łatwo dostosować do swoich zadań, a ogromna społeczność użytkowników zawsze pomoże znaleźć rozwiązanie każdego problemu.

RDBMS pozostaje podstawą dla wielu aplikacji biznesowych i platform analitycznych (Rys. 3.1-6), które umożliwiają firmom wydajne przechowywanie, przetwarzanie i analizowanie danych - a tym samym podejmowanie świadomych i terminowych decyzji.



Rys. 3.1-6 Ankieta przeprowadzona wśród programistów na StackOverFlow (największym forum IT) dotycząca baz danych, z których korzystali w zeszłym roku i z których chcą korzystać w przyszłym roku (RDBMS są zaznaczone na niebiesko) (na podstawie [50]).

RDBMS zapewnia niezawodność, spójność danych, obsługę transakcji i wykorzystuje potężny język zapytań - SQL (Structured Query Language), który jest często używany w analityce i umożliwia łatwe uzyskiwanie, modyfikowanie i analizowanie informacji przechowywanych w bazach danych. SQL jest głównym narzędziem do pracy z danymi w systemach relacyjnych.

SQL - zapytania w bazach danych i nowe trendy

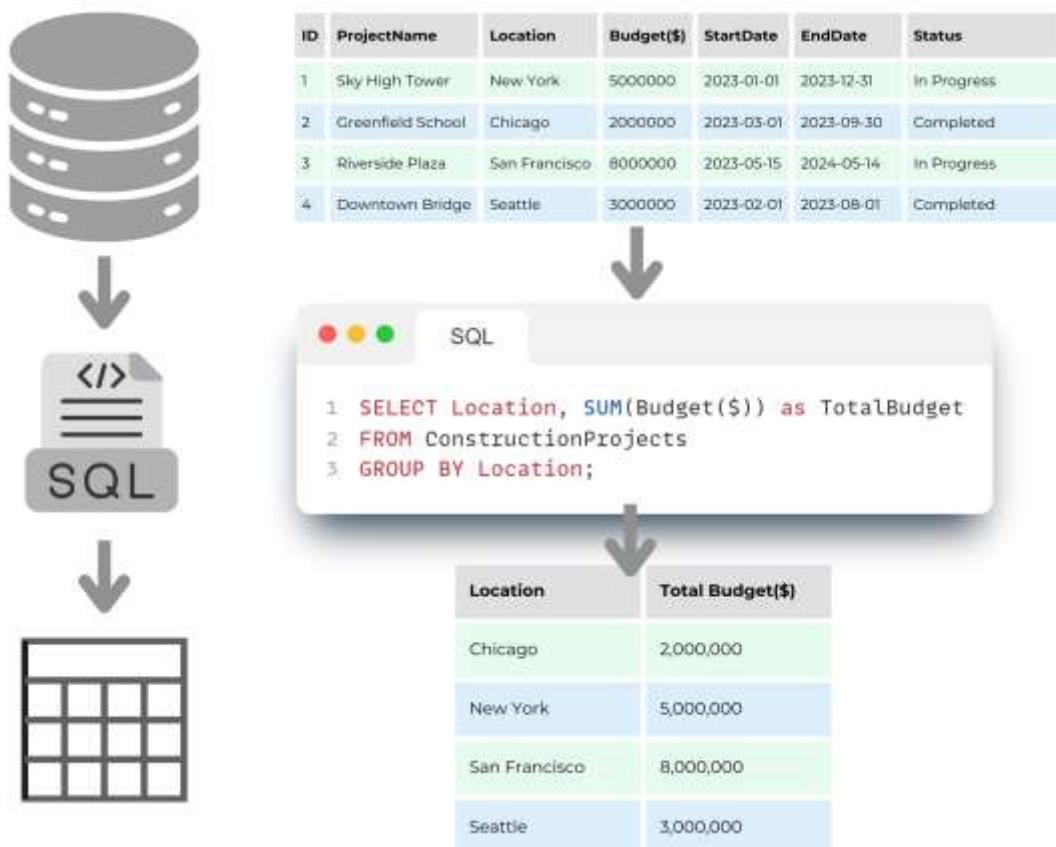
Główną zaletą języka SQL, często używanego w relacyjnych bazach danych, w porównaniu z innymi rodzajami zarządzania informacjami (na przykład za pomocą klasycznych arkuszy kalkulacyjnych Excel) jest obsługa bardzo dużych wolumenów baz danych przy dużej szybkości przetwarzania zapytań.

Structured Query Language (SQL) to specjalistyczny język programowania przeznaczony do przechowywania, przetwarzania i analizowania informacji w relacyjnych bazach danych. SQL służy do tworzenia, zarządzania i uzyskiwania dostępu do danych, umożliwiając wydajne wyszukiwanie, filtrowanie, łączenie i agregowanie informacji. Służy jako kluczowe narzędzie dostępu do danych, zapewniając wygodny i sformalizowany sposób interakcji z magazynami informacji.

Ewolucja systemów SEQUEL-SQL przechodzi przez znaczące produkty i firmy, takie jak Oracle, IBM DB2, Microsoft SQL Server, SAP, PostgreSQL i MySQL, a jej kulminacją jest pojawienie się SQLite i MariaDB [51]. SQL zapewnia możliwości arkusza kalkulacyjnego, których nie można znaleźć w programie Excel, dzięki czemu manipulowanie danymi jest bardziej skalowalne, bezpieczne i łatwe do zautomatyzowania:

- **Tworzenie i zarządzanie strukturami danych (DDL):** W SQL można tworzyć, modyfikować i usuwać tabele w bazie danych, ustanawiać powiązania między nimi i definiować struktury przechowywania danych. Excel, z drugiej strony, działa z ustalonymi arkuszami i komórkami, bez jasno zdefiniowanych relacji między arkuszami i zestawami danych.

- **Manipulacja danymi (DML):** SQL umożliwia masowe dodawanie, modyfikowanie, usuwanie i pobieranie danych z dużą prędkością poprzez wykonywanie złożonych zapytań z filtrowaniem, sortowaniem i łączeniem tabel (Rys. 3.1-7). W programie Excel przetwarzanie dużych ilości informacji wymaga ręcznych działań lub specjalnych makr, co spowalnia proces i zwiększa prawdopodobieństwo wystąpienia błędów.
- **Kontrola dostępu (DCL):** SQL umożliwia zróżnicowanie praw dostępu do danych dla różnych użytkowników, ograniczając możliwość edycji lub przeglądania informacji. Z kolei w Excel dostęp jest albo współdzielony (podczas przesyłania pliku), albo wymaga złożonych ustawień z udostępnianiem uprawnień za pośrednictwem usług w chmurze.



Rys. 3.1-7 Przykład DML w SQL: szybkie przetwarzanie, grupowanie i agregacja za pomocą kilku wierszy kodu do automatycznego przetwarzania danych.

Excel ułatwia pracę z danymi dzięki swojej wizualnej i intuicyjnej strukturze. Jednak wraz ze wzrostem ilości danych wydajność programu Excel spada. Excel napotyka również ograniczenia dotyczące ilości danych, które może przechowywać - maksymalnie milion wierszy - a wydajność spada na dłużej przed osiągnięciem tego limitu. Tak więc, podczas gdy Excel wydaje się preferowany do wizualizacji i manipulowania niewielkimi ilościami danych, SQL lepiej nadaje się do radzenia sobie z dużymi zestawami danych.

Kolejnym etapem w rozwoju danych strukturalnych było pojawienie się kolumnowych baz danych (Columnar Databases), które stanowią alternatywę dla tradycyjnych relacyjnych baz danych, zwłaszcza w przypadku znacznie większych wolumenów danych i obliczeń analitycznych. W przeciwieństwie do wierszowych baz danych, w których dane są przechowywane wiersz po wierszu, kolumnowe bazy danych

rejestrują informacje według kolumn. W porównaniu do klasycznych baz danych, pozwala to na:

- Zmniejszenie przestrzeni dyskowej poprzez wydajną kompresję jednolitych danych w kolumnach.
- Przyspieszenie zapytań analitycznych, ponieważ odczytywane są tylko wymagane kolumny, a nie cała tabela.
- Optymalizacja Big Data i hurtowni danych, np. architektura Data Lakehouse.

Więcej o kolumnowych bazach danych, Pandas DataFrame, Apache Parquet, HDF5 także o tworzeniu na ich podstawie magazynów Big Data do celów analizy i przetwarzania danych powiemy w kolejnych rozdziałach tej książki - "DataFrame: uniwersalny tabelaryczny format danych" oraz "Formaty przechowywania danych i praca z Apache Parquet: DWH - hurtownie danych i architektura Data Lakehouse".

Dane niestrukturyzowane

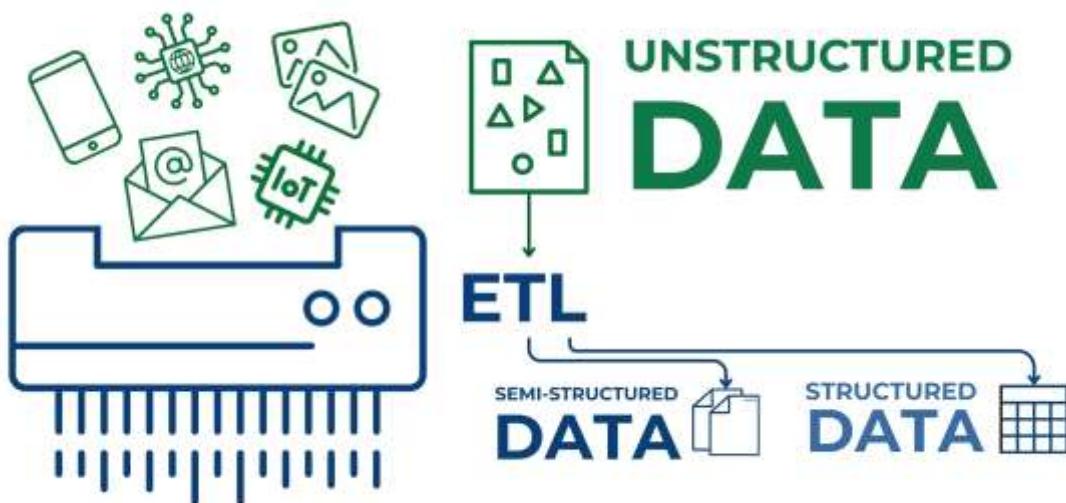
Chociaż większość danych wykorzystywanych w aplikacjach i systemach informatycznych ma postać ustrukturyzowaną, większość informacji generowanych w budownictwie ma postać danych niestrukturyzowanych - obrazów, filmów, dokumentów tekstowych, nagrań audio i innych form treści. Jest to szczególnie prawdziwe na etapie budowy, eksploatacji i nadzoru technicznego, gdzie dominują informacje wizualne i tekstowe.

Dane niestrukturyzowane to informacje, które nie mają wstępnie zdefiniowanego modelu lub struktury, nie są zorganizowane w tradycyjne wiersze i kolumny, jak w bazach danych lub tabelach.

Ogólnie rzecz biorąc, dane niestrukturyzowane można podzielić na dwie kategorie:

- Dane niestrukturyzowane generowane przez człowieka, które obejmują różne rodzaje treści generowanych przez człowieka: dokumenty tekstowe, wiadomości e-mail, obrazy, filmy itp.
- Niestrukturyzowane dane generowane maszynowo są tworzone przez urządzenia i czujniki: obejmują one pliki dziennika, dane GPS, wyniki Internetu rzeczy (IoT) i inne informacje telemetryczne, na przykład z placu budowy.

W przeciwieństwie do danych ustrukturyzowanych, które są wygodnie zorganizowane w tabelach i bazach danych, dane niestrukturyzowane wymagają dodatkowych etapów przetwarzania przed ich integracją z systemami informatycznymi (rys. 3.1-8). Wykorzystanie technologii do zautomatyzowanego gromadzenia, analizy i przekształcania takich danych otwiera nowe możliwości poprawy wydajności budowy, zmniejszenia liczby błędów i zminimalizowania wpływu czynnika ludzkiego.



Rys. 3.1-8 Przetwarzanie niestrukturyzowanych danych rozpoczyna się od ich przekształcenia w dane częściowo ustrukturyzowane i ustrukturyzowane.

Dane niestrukturyzowane stanowią do 80% wszystkich informacji [52] napotykanych przez profesjonalistów w firmach, dlatego w kolejnych rozdziałach książki omówimy szczegółowo ich rodzaje i przetwarzanie wraz z przykładami.

Dla ułatwienia dyskusji dane tekstowe zostały podzielone na osobne kategorie. Chociaż są one rodzajem raczej niestrukturyzowanych danych, ich znaczenie i powszechność w branży budowlanej wymagają szczególnej uwagi.

Dane tekstowe: między chaosem niestrukturyzowanym a chaosem ustrukturyzowanym

Dane tekstowe w branży budowlanej obejmują szeroki zakres formatów i rodzajów informacji, od dokumentów papierowych po nieformalne metody komunikacji, takie jak listy, rozmowy, korespondencja robocza i ustne spotkania na placu budowy. Wszystkie te dane tekstowe niosą ze sobą ważne informacje dla zarządzania projektami budowlanymi, od szczegółów decyzji projektowych i zmian w planach po dyskusje na temat kwestii bezpieczeństwa i negocjacje z wykonawcami i klientami (Rys. 3.1-9).



Rys. 3.1-9 Dane tekstowe, jeden z najpopularniejszych typów informacji wykorzystywanych w komunikacji pomiędzy uczestnikami projektu.

Informacje tekstowe mogą być zarówno sformalizowane, jak i niestrukturyzowane. Sformalizowane dane obejmują dokumenty Word (.doc,.docx), PDF, a także pliki tekstowe protokołów spotkań (.txt). Dane niesformalizowane obejmują korespondencję za pośrednictwem komunikatorów i poczty, transkrypcje spotkań (Teams, Zoom, Google Meet) oraz nagrania audio dyskusji (.mp3,.wav), które wymagają konwersji na tekst.

O ile jednak dokumenty pisemne, takie jak formalne wnioski, warunki umów i wiadomości e-mail zazwyczaj mają już pewną strukturę, o tyle komunikacja ustna i korespondencja robocza często pozostają niestrukturyzowane, co utrudnia ich analizę i integrację z systemami zarządzania projektami.

Kluczem do efektywnego zarządzania danymi tekstowymi jest ich konwersja do formatu ustrukturyzowanego. Pozwala to na automatyczną integrację przetworzonych informacji z istniejącymi systemami, które już pracują z danymi strukturalnymi.



Rys. 3.1-10 Konwersja treści tekstuowej na dane strukturalne.

Aby efektywnie wykorzystać informacje tekstowe, muszą one zostać automatycznie przekształcone w ustrukturyzowaną formę (Rys. 3.1-10). Proces ten obejmuje zazwyczaj kilka etapów:

- **Rozpoznawanie tekstu (OCR)** - konwertowanie obrazów dokumentów i rysunków do formatu nadającego się do odczytu maszynowego.
- **Analiza tekstu (NLP)** - automatyczna identyfikacja kluczowych parametrów (daty, kwoty i liczby istotne dla projektu).
- **Klasyfikacja danych** - kategoryzacja informacji (finanse, logistyka, zarządzanie ryzykiem).

Po rozpoznaniu i sklasyfikowaniu już ustrukturyzowane dane można zintegrować z bazami danych i

wykorzystać w zautomatyzowanych systemach raportowania i zarządzania.

Dane częściowo ustrukturyzowane i luźno ustrukturyzowane

Dane częściowo ustrukturyzowane zawierają pewien poziom organizacji, ale nie mają ścisłego schematu ani struktury. Chociaż takie informacje zawierają ustrukturyzowane elementy (np. daty, nazwiska pracowników i listy wykonanych zadań), format prezentacji może się znacznie różnić w zależności od projektu, a nawet od jednego pracownika do drugiego. Przykładami takich danych są dzienniki czasu, raporty z postępów i harmonogramy, które mogą być prezentowane w różnych formatach.

Dane częściowo ustrukturyzowane są łatwiejsze do analizy niż dane nieustrukturyzowane, ale wymagają dodatkowego przetwarzania w celu integracji ze standardowymi systemami zarządzania projektami.

Praca z częściowo ustrukturyzowanymi danymi charakteryzującymi się stale zmieniającą się strukturą stanowi poważne wyzwanie. Wynika to z faktu, że zmienność struktury danych wymaga odrębnego, indywidualnego podejścia do przetwarzania i analizy każdego źródła danych częściowo ustrukturyzowanych.

Podczas gdy przetwarzanie nieustrukturyzowanych danych wymaga wiele wysiłku, przetwarzanie danych częściowo ustrukturyzowanych może być wykonywane przy użyciu stosunkowo prostych metod i narzędzi.

Słabo ustrukturyzowane dane to bardziej ogólny termin opisujący dane o minimalnej lub niekompletnej strukturze. Najczęściej są to dokumenty tekstowe, czaty, e-maile, w których znajdują się pewne metadane (np. data, nadawca), ale większość informacji jest prezentowana w chaotyczny sposób.

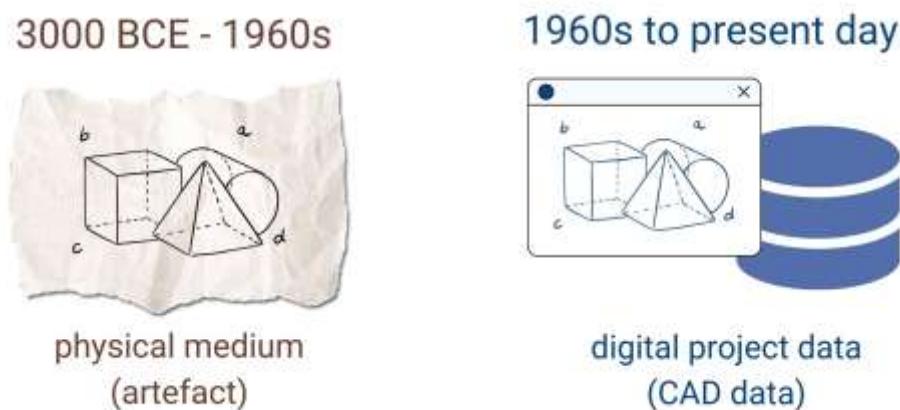
W budownictwie luźno ustrukturyzowane dane występują w różnych procesach. Na przykład mogą one obejmować:

- Szacunki i wyceny - tabele z danymi dotyczącymi materiałów, ilości i kosztów, ale bez jednolitego formatu.
- Rysunki i schematy inżynieryjne - pliki w formacie PDF lub DWG, zawierające adnotacje tekstowe i metadane, ale bez ścisłe ustalonej struktury.
- Harmonogramy pracy - dane z MS Project, Primavera P6 lub innych systemów, które mogą mieć inną strukturę eksportu.
- CAD (BIM -models) - zawierają elementy konstrukcji, ale reprezentacja danych zależy od oprogramowania i standardu projektu.

Dane geometryczne, generowane przez systemy CAD, mogą być kategoryzowane w taki sam sposób, jak dane częściowo ustrukturyzowane. Będziemy jednak klasyfikować geometryczne dane CAD (BIM) jako oddzielny typ danych, ponieważ podobnie jak dane tekstowe, często mogą być traktowane jako oddzielny typ danych w procesach firmy.

Dane geometryczne i ich zastosowanie

Podczas gdy metadane dotyczące elementów projektu są prawie zawsze przechowywane w formie tabel, ustrukturyzowanych lub luźno ustrukturyzowanych formatów, dane geometryczne elementów projektu w większości przypadków są tworzone przy użyciu specjalnych narzędzi CAD (rys. 3.1-11), umożliwiających szczegółową wizualizację elementów projektu jako zestawu linii (2D) lub brył geometrycznych (3D).



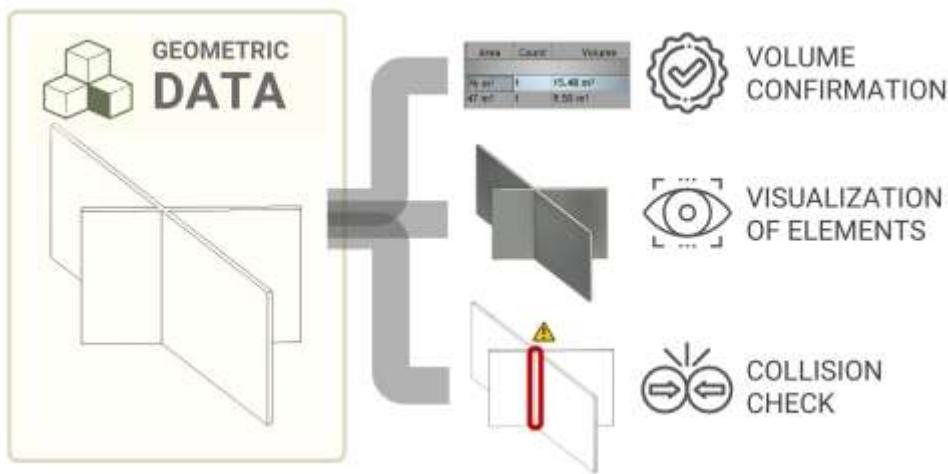
Rys. 3.1-11 Narzędzia CAD pomogły przenieść informacje geometryczne z nośnika fizycznego do postaci bazy danych.

Podczas pracy z danymi geometrycznymi w budownictwie i architekturze można zidentyfikować trzy główne zastosowania danych geometrycznych (Rys. 3.1-12):

- **Potwierdzenie objętości:** dane geometryczne, generowane w programach CAD (BIM) przy użyciu specjalnych jąder geometrycznych, są wymagane do automatycznego i dokładnego określenia objętości i wymiarów elementów projektu. Dane te obejmują automatycznie obliczane powierzchnie, objętości, długości i inne ważne atrybuty wymagane do planowania, budżetowania i zamawiania zasobów i materiałów
- **Wizualizacja projektu:** w przypadku jakichkolwiek zmian w projekcie, wizualizacja elementów pozwala na automatyczne generowanie zaktualizowanych rysunków w różnych płaszczyznach. Wizualizacja projektu na początkowych etapach pomaga przyspieszyć zrozumienie między wszystkimi uczestnikami, aby zaoszczędzić czas i zasoby podczas procesu budowy.
- **Sprawdzanie kolizji:** W złożonych projektach budowlanych i inżynierijnych, w których interakcja wielu kategorii elementów (np. rur i ścian) bez "konfliktów geometrycznych" ma kluczowe znaczenie, sprawdzanie kolizji odgrywa kluczową rolę. Korzystanie z oprogramowania do wykrywania kolizji pozwala proaktywnie identyfikować potencjalne konflikty geometryczne między elementami projektu, zapobiegając kosztownym błędom podczas procesu budowy.

Od samego początku istnienia inżynierskich biur projektowych, od czasu budowy pierwszych złożonych konstrukcji, inżynierowie budownictwa dostarczali informacje geometryczne w postaci rysunków, linii i płaskich elementów geometrycznych (na papirusie, papierze "A0" lub w formatach DWG, PDF, PLT), na podstawie których majstrowie i kosztorysanci (rys. 3.1-11), przez ostatnie tysiąclecia, za pomocą linijek i kątomierzy, zbierali atrybutowe objętości lub ilości elementów i grup elementów (rys. 3.1-11), przez ostatnie tysiąclecia, za pomocą linijek i transporterów, zbierali atrybutywne objętości lub ilości elementów i grup

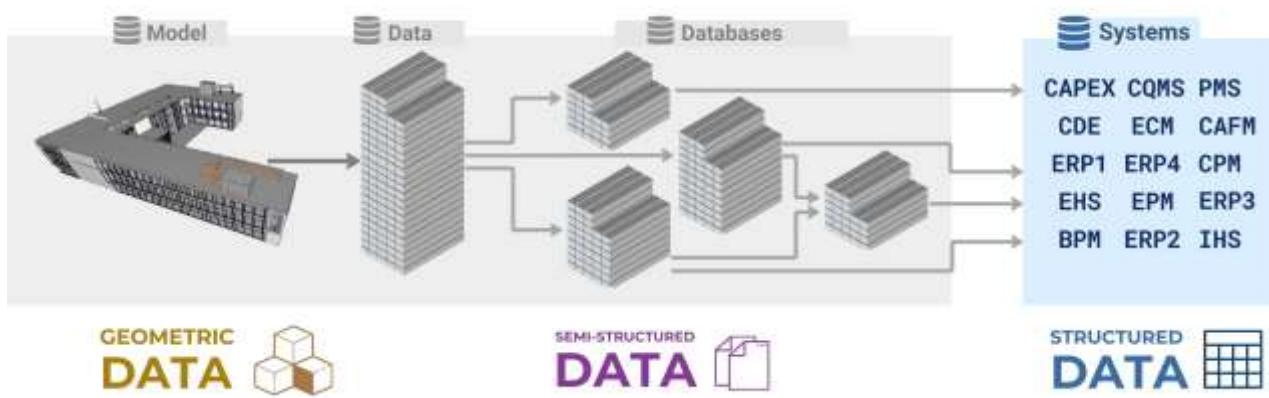
elementów.



Rys. 3.1-12 Geometria jest podstawą do uzyskania parametrów objętościowych elementów, które są następnie wykorzystywane do obliczania kosztów i harmonogramu projektu.

Obecnie to ręczne i czasochłonne zadanie jest rozwiązywane przez pełną automatyzację dzięki pojawiению się modelowania wolumetrycznego w nowoczesnych narzędziach CAD (BIM), które pozwala automatycznie, za pomocą specjalnego jądra geometrycznego, uzyskać atrybuty wolumetryczne dowolnego elementu bez konieczności ręcznego obliczania parametrów wolumetrycznych.

Nowoczesne narzędzia CAD umożliwiają również klasyfikowanie i kategoryzowanie elementów projektu, dzięki czemu można przesyłać tabele specyfikacji z bazy danych projektu do wykorzystania w różnych systemach, takich jak szacowanie kosztów, planowanie lub obliczanie CO₂ (rys. 3.1-13). Uzyskiwanie specyfikacji, tabele QTO i ilości oraz praktyczne przykłady omówimy w rozdziale "Uzyskiwanie ilości i kwantyfikacja".



Rys. 3.1-13 Narzędzia CAD (BIM) przechowują dane w bazach danych, które są zaprojektowane do integracji i interakcji z innymi systemami.

Ze względu na zamknięty charakter baz danych i formatów wykorzystywanych w środowisku CAD, dane

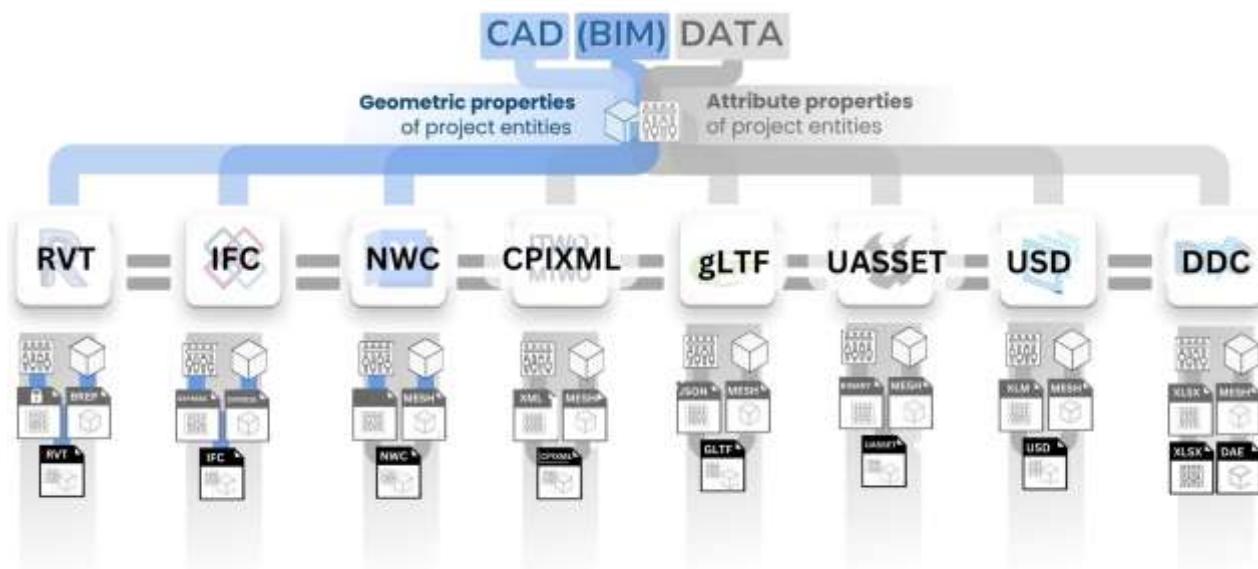
geometryczne tworzone w rozwiązańach CAD stały się w rzeczywistości odrębnym rodzajem informacji. Łączą one w sobie zarówno geometrię elementów, jak i metainformacje (ustrukturyzowane lub częściowo ustrukturyzowane), zamknięte w wyspecjalizowanych plikach i formatach.

Dane CAD: od projektu do przechowywania danych

Nowoczesne systemy CAD i BIM przechowują dane we własnych, często zastrzeżonych formatach: DWG, DXF, RVT, DGN, PLN i innych. Formaty te obsługują zarówno dwu-, jak i trójwymiarowe reprezentacje obiektów, zachowując nie tylko geometrię, ale także atrybuty związane z obiektami. Oto najpopularniejsze z nich:

- **DWG** to binarny format pliku używany do przechowywania dwuwymiarowych (i rzadziej trójwymiarowych) danych projektowych i metadanych.
- **DXF** to format tekstowy służący do wymiany rysunków 2D i 3D między systemami CAD. Zawiera geometrię, warstwy i dane atrybutów, obsługuje zarówno reprezentację ASCII, jak i binarną.
- **RVT** to format binarny do przechowywania modeli CAD, w tym geometrii 3D, atrybutów elementów, relacji i parametrów projektowych.
- **IFC** to otwarty format tekstowy służący do wymiany danych budowlanych między systemami CAD (BIM). Obejmuje on geometrię, właściwości obiektów i informacje o ich relacjach.

Oprócz tego używane są inne formaty: PLN, DB1, SVF, NWC, CPIXML, BLEND, BX3, USD, XLSX, DAE. Chociaż różnią się one celem i poziomem otwartości (Rys. 3.1-14), wszystkie mogą reprezentować ten sam model informacyjny projektu w różnych formach. W złożonych projektach formaty te są często używane równolegle, od rysowania do koordynowania modeli projektów.



Rys. 3.1-14 Popularne formaty zapisu CAD opisują geometrię za pomocą parametrów BREP lub MESH, uzupełnionych danymi atrybutów.

Wszystkie powyższe formaty umożliwiają przechowywanie danych o każdym elemencie projektu budowlanego i wszystkie powyższe formaty zawierają dwa kluczowe typy danych:

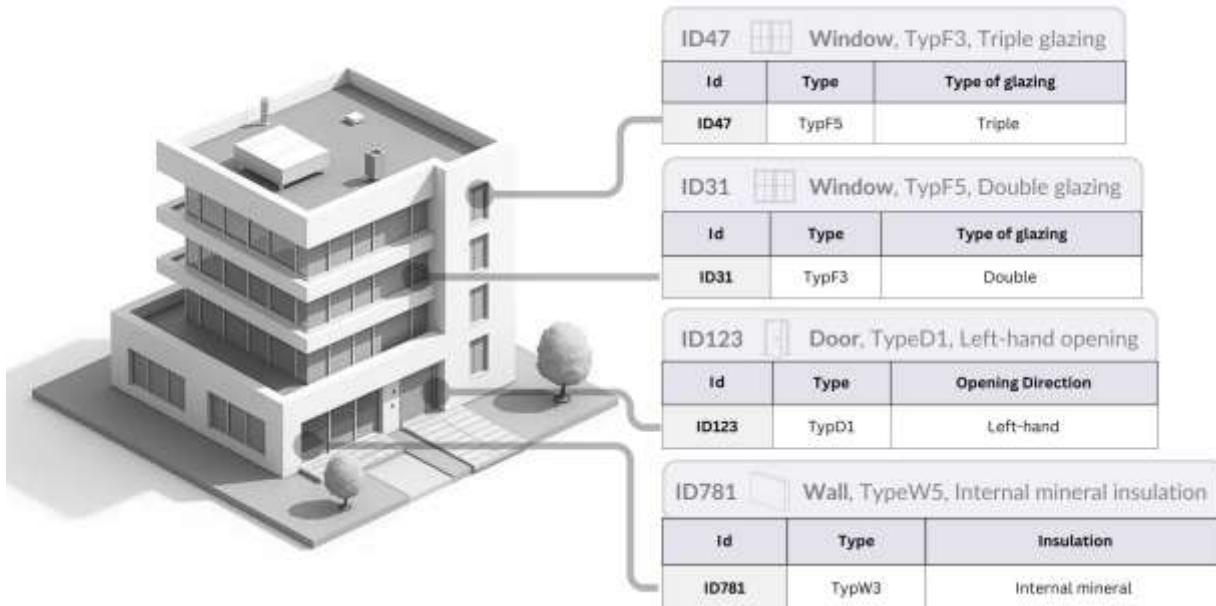
- **Parametry geometryczne** - opisują kształt, położenie i wymiary obiektu. Geometria i jej wykorzystanie zostaną szczegółowo omówione w szóstej części książki poświęconej rozwiązaniom CAD (BIM);
- **Właściwości atrybutów** - zawierają różne informacje: materiały, typy elementów, charakterystyki techniczne, unikalne identyfikatory i inne właściwości, które mogą mieć elementy projektu.

Dane atrybutów mają szczególne znaczenie w nowoczesnych projektach, ponieważ definiują charakterystykę operacyjną obiektów, umożliwiają obliczenia inżynieryjne i kosztorysowe oraz zapewniają kompleksową interakcję między uczestnikami projektowania, budowy i eksploatacji. Na przykład:

- W przypadku okien i drzwi: rodzaj konstrukcji, rodzaj oszklenia, kierunek otwierania (rys. 3.2-1).
- W przypadku ścian rejestrowane są informacje o materiałach, izolacji termicznej i właściwościach akustycznych.
- W przypadku systemów inżynieryjnych przechowywane są parametry rurociągów, kanałów, tras kablowych i ich połączeń.

Parametry te mogą być przechowywane zarówno w samych plikach CAD-(BIM-) jak i w zewnętrznych bazach danych - w wyniku eksportu, konwersji lub bezpośredniego dostępu do wewnętrznych struktur CAD za pomocą narzędzi inżynierii odwrotnej. Takie podejście ułatwia integrację informacji projektowych z innymi systemami i platformami korporacyjnymi

Inżynieria odwrotna w kontekście CAD (BIM) to proces wyodrębniania i analizowania wewnętrznej struktury modelu cyfrowego w celu odtworzenia jego logiki, struktury danych i zależności bez dostępu do oryginalnych algorytmów lub dokumentacji.

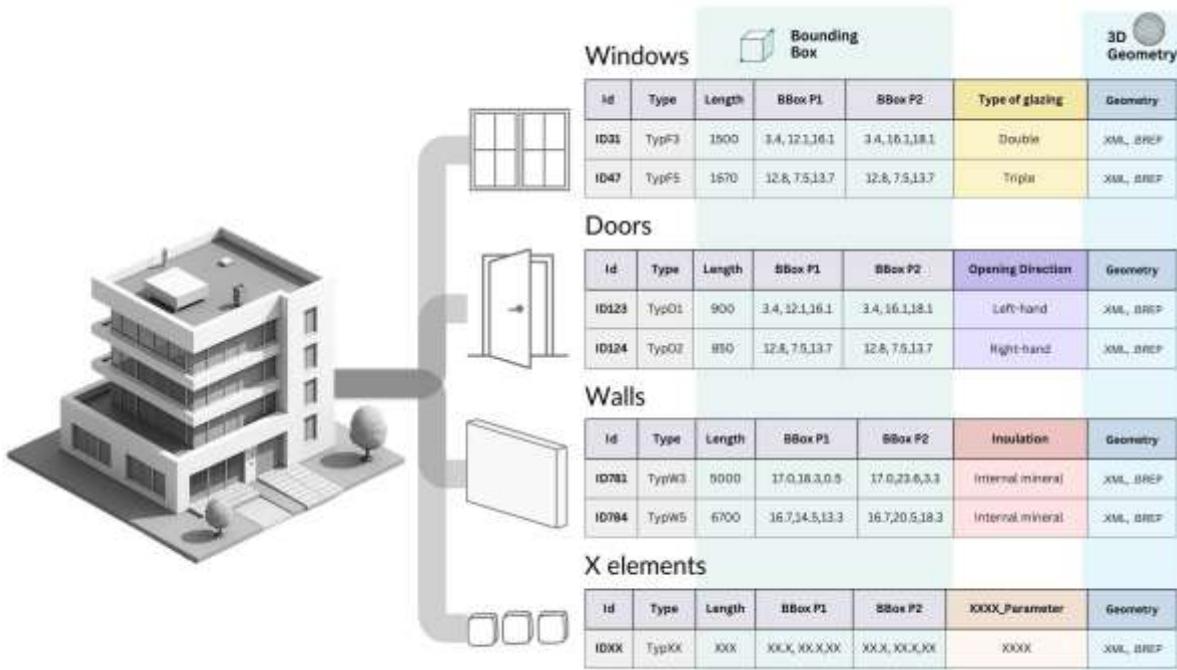


Rys. 3.1-15 Element projektu, oprócz opisu geometrii parametrycznej lub wielokątnej, zawiera informacje o parametrach i właściwościach elementów.

W rezultacie wokół każdego elementu tworzony jest unikalny zestaw parametrów i właściwości, obejmujący

zarówno unikalne cechy każdego obiektu (np. identyfikator i wymiary), jak i wspólne atrybuty dla grup elementów. Pozwala to nie tylko analizować poszczególne elementy-podmioty projektu, ale także łączyć je w logiczne grupy, które mogą być następnie wykorzystywane przez innych specjalistów do ich zadań i obliczeń w systemach i bazach danych.

Encja to konkretny lub abstrakcyjny obiekt świata rzeczywistego, który można jednoznacznie zidentyfikować, opisać i przedstawić w postaci danych.

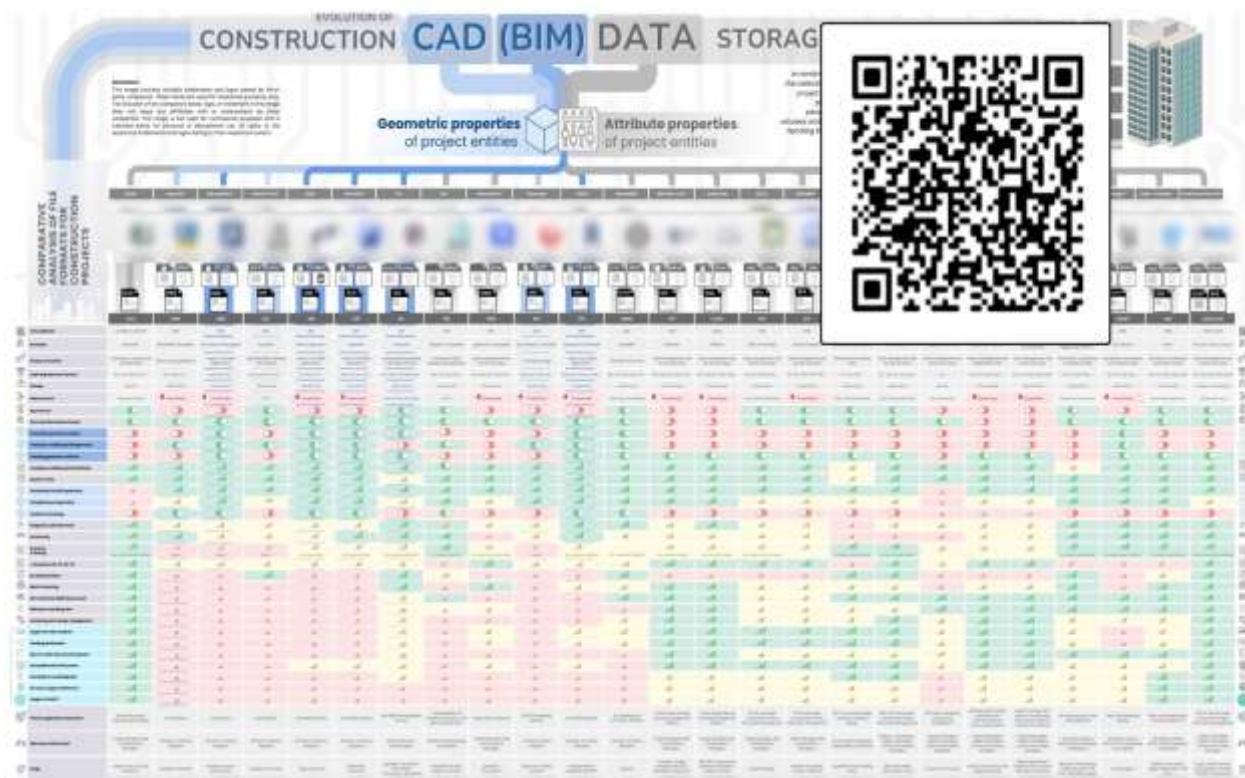


Rys. 3.1-16 Każdy element projektu zawiera atrybuty, które są wprowadzane przez projektanta lub obliczane w programie CAD.

W ciągu ostatnich dziesięcioleci branża budowlana opracowała wiele nowych formatów CAD (BIM), które upraszczają tworzenie, przechowywanie i przesyłanie danych. Formaty te mogą być zamknięte lub otwarte, tabelaryczne, parametryczne lub graficzne. Jednak ich różnorodność i fragmentacja znacznie komplikują zarządzanie danymi na wszystkich etapach cyklu życia projektu. Tabela porównawcza głównych formatów wykorzystywanych do wymiany informacji w budownictwie została przedstawiona na rysunku 3.1-17 (pełna wersja dostępna za pomocą kodu QR).

Aby rozwiązać problemy interoperacyjności i dostępu do danych CAD, zaangażowani są menedżerowie BIM (BIM) i koordynatorzy, których zadaniem jest kontrola eksportu, sprawdzanie jakości danych i integracja części danych CAD (BIM) z innymi systemami.

Jednak ze względu na zamknięty charakter i złożoność formatów, trudno jest zautomatyzować ten proces, co zmusza specjalistów do wykonywania wielu operacji ręcznie, bez możliwości budowania pełnoprawnych procesów przetwarzania danych in-line (pipeline).



The diagram illustrates the evolution of construction CAD (BIM) data storage. It compares various data formats used in project entities, categorized into geometric properties and attribute properties. The grid shows the relationship between different formats, such as DXF, DWG, IFC, and others, highlighting how they have evolved from separate entities to integrated components within a single BIM model.

Rys. 3.1-17 Tabela porównująca główne formaty danych, w których przechowywane są informacje o elementach projektu [53].

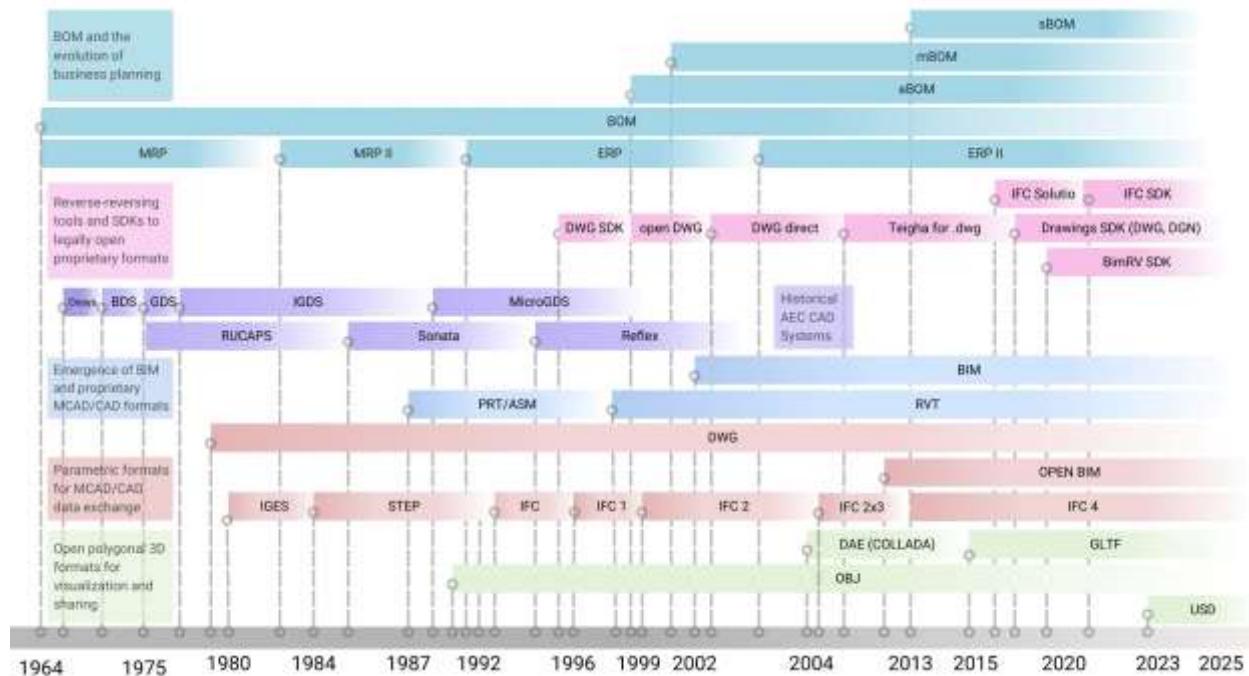
Aby zrozumieć, dlaczego istnieje tak wiele różnych formatów danych i dlaczego większość z nich jest zamknięta, ważne jest, aby zagłębić się w procesy zachodzące wewnętrz programów CAD (BIM), które zostaną szczegółowo zbadane w szóstej części książki.

Dodatkowa warstwa informacji dodana do geometrii została wprowadzona przez twórców systemów CAD w postaci koncepcji BIM (Building Information Modeling), terminu marketingowego aktywnie promowanego w branży budowlanej od 2002 roku [54].

Pojawienie się koncepcji BIM (BOM) i wykorzystanie CAD w procesach

Koncepcja modelowania informacji o budynku (BIM), po raz pierwszy nakreślona w dokumencie BIM Whitepaper z 2002 roku [54], wywodzi się z inicjatyw marketingowych producentów oprogramowania CAD. Wyłoniła się z inicjatyw marketingowych twórców oprogramowania CAD i była próbą dostosowania zasad już dobrze ugruntowanych w inżynierii mechanicznej do potrzeb branży budowlanej.

Inspiracją dla BIM była koncepcja BOM (Bill of Materials) - zestawienie materiałów - która była szeroko stosowana w przemyśle od późnych lat 80-tych. W inżynierii mechanicznej BOM umożliwił połączenie danych z systemów CAD z systemami PDM (Product Data Management), PLM (Product Lifecycle Management) i ERP, zapewniając holistyczne zarządzanie informacjami inżynierijnymi w całym cyklu życia produktu (rys. 3.1-8).



Rys. 3.1-18 Ewolucja specyfikacji (BOM), modelowania informacji (BIM) i formatów cyfrowych w branży inżynierijno-budowlanej.

Nowoczesny rozwój koncepcji BOM doprowadził do powstania rozszerzonej struktury - XBOM (Extended BOM), która obejmuje nie tylko skład produktu, ale także scenariusze behawioralne, wymagania operacyjne, parametry zrównoważonego rozwoju i dane do analizy predykcyjnej. XBOM zasadniczo spełnia tę samą rolę, co BIM w budownictwie: oba podejścia mają na celu przekształcenie modelu cyfrowego w Jedno Źródło Prawdy dla wszystkich uczestników projektu przez cały cykl życia projektu.

Kluczowym kamieniem milowym w pojawienniu się BOM w budownictwie było wprowadzenie pierwszego parametrycznego systemu CAD (MCAD) specjalnie dostosowanego do branży budowlanej w 2002 roku. Został on opracowany przez zespół, który wcześniej stworzył Pro-E®, rewolucyjny system MCAD dla inżynierii mechanicznej, który pojawił się pod koniec lat 80. i stał się standardem branżowym [55].

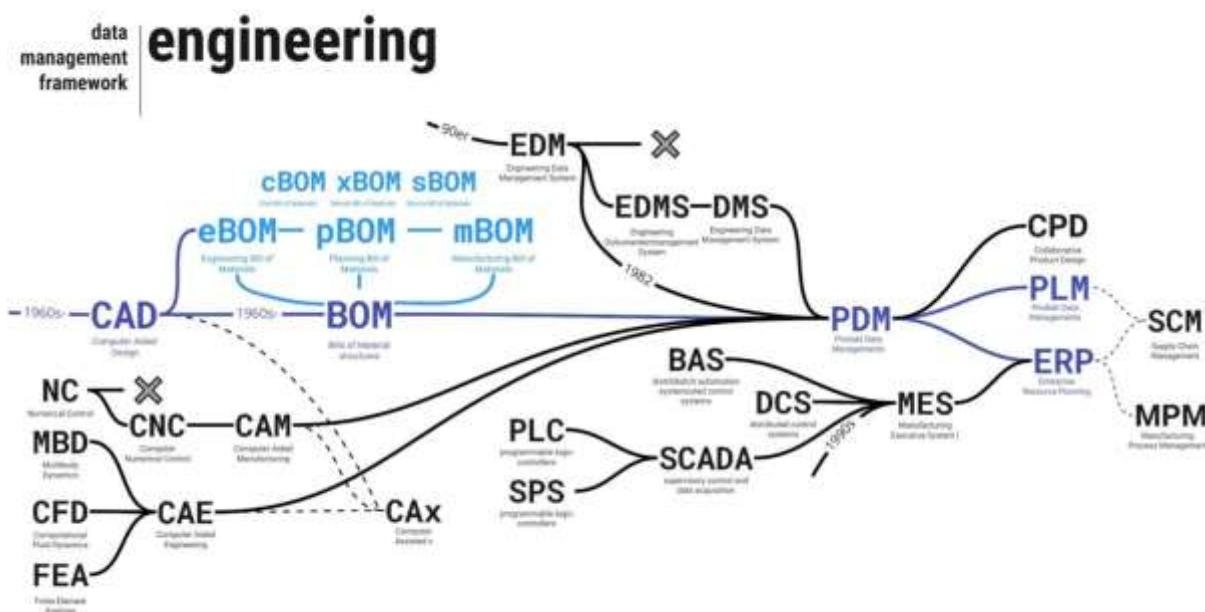
Już pod koniec lat 80-tych dążono do wyeliminowania ograniczeń [56] istniejących wówczas programów CAD. Głównym celem było zmniejszenie pracochłonności wprowadzania zmian parametrów elementów konstrukcyjnych oraz umożliwienie aktualizacji modelu na podstawie danych spoza programów CAD za pośrednictwem bazy danych [57]. Najważniejszą rolę w tym miała odgrywać parametryzacja: automatyczne pobieranie charakterystyk z bazy danych i wykorzystywanie ich do aktualizacji modelu wewnętrz systemów CAD.

Pro-E i leżąca u jego podstaw koncepcja elementarnego modelowania parametrycznego c BOM miały znaczący wpływ na rozwój rynku CAD - i MCAD - [58]. Przez 25 lat model ten był obecny w branży, a wiele nowoczesnych systemów stało się jego koncepcyjnymi następcami.

Celem jest stworzenie systemu, który jest wystarczająco elastyczny, aby zachęcić inżyniera do łatwego rozważania różnych projektów. Koszt wprowadzania zmian w projekcie powinien być jak najbliższy zera. Tradycyjne oprogramowanie CAD / CAM nierealistycznie ogranicza wprowadzanie niedrogich zmian tylko na samym początku procesu projektowania [59].

- Samuel Heisenberg, założyciel Parametric Technology Corporation®, twórca MCAD - produktu Pro-E i nauczyciel twórcy produktu CAD wykorzystującego format RVT

W inżynierii mechanicznej systemy PDM, PLM, MRP i ERP stały się kluczowymi platformami. Odgrywają one centralną rolę w zarządzaniu danymi i procesami, gromadząc informacje z systemów CAx (CAD, CAM, CAE) i organizując działania projektowe w oparciu o strukturę produktu (BOM: eBOM, pBOM, mBOM) (rys. 3.1-18). Taka integracja zmniejsza liczbę błędów, pozwala uniknąć powielania danych i zapewnia kompleksową identyfikowalność od projektu do produkcji.



Rys. 3.1-19 Historycznie, BOM pojawił się w latach 60-tych XX wieku jako sposób na ustrukturyzowanie danych z systemów CAx i przekazanie ich do systemów sterowania.

Zakup przez jednego z wiodących dostawców rozwiązania CAD opracowanego przez były zespół Pro-E i opartego na podejściu BOM został niemal natychmiast oznaczony publikacją serii BIM Whitepaper (2002-2003)[60][61]. Już od połowy XXI wieku koncepcja BIM zaczęła być aktywnie promowana w branży budowlanej, co znacznie zwiększyło zainteresowanie oprogramowaniem parametrycznym. Popularność wzrosła tak szybko, że widełki konstrukcyjne inżynierii mechanicznej Pro-E - parametryczny CAD promowany przez tego dostawcę - faktycznie wyparł konkurentów w segmencie projektowania architektonicznego i konstrukcyjnego (rys. 3.1-20). Na początku XXI wieku de facto skonsolidował globalną dominację na rynku BIM (CAD) [62].



Rys. 3.1-20 Popularność zapytań w wyszukiwarce Google (RVT kontra IFC): parametryczny CAD stworzony przez były zespół Pro-E z obsługą BOM -BIM zyskał popularność w prawie większości krajów świata.

W ciągu ostatnich 20 lat skrót BIM zyskał wiele interpretacji, z których wiele znaczeń ma swoje korzenie w początkowych koncepcjach marketingowych, które pojawiły się na początku XXI wieku. Norma ISO 19650, która odegrała ważną rolę w popularyzacji tego terminu, faktycznie zapewniła BIM status "naukowego" podejścia do zarządzania informacją. Jednak w samym tekście normy, który poświęcony jest zarządzaniu danymi w całym cyklu życia obiektów z wykorzystaniem BIM, skrót BIM jest wspomniany, ale nigdy nie został jasno zdefiniowany

Oryginalna strona internetowa dostawcy, która opublikowała serię białych dokumentów na temat BIM w 2002 r.[60] i 2003 r.[61], w rzeczywistości powiełała materiały marketingowe dotyczące koncepcji BOM (Bills of Materials) i PLM (Product Lifecycle Management), które były wcześniej używane w oprogramowaniu inżynierii mechanicznej Pro-E w latach 90-tych [63].

Modelowanie Informacji o Budynku, innowacyjne nowe podejście do projektowania, budowy i zarządzania budynkami wprowadzone przez..... [nazwa firmy dostawcy CAD] w 2002 roku, zmieniło sposób myślenia profesjonalistów z branży na całym świecie o tym, w jaki sposób można zastosować technologię do projektowania, budowy i zarządzania budynkami.

- Biała Księga BIM, 2003 [61]

Te wczesne publikacje łączyły BIM bezpośrednio z koncepcją scentralizowanej, zintegrowanej bazy danych. Jak stwierdzono w dokumencie Whitepaper z 2003 r., BIM to zarządzanie informacjami o budynku, w którym wszystkie aktualizacje mają miejsce w jednym repozytorium, zapewniając synchronizację wszystkich rysunków, cięć i specyfikacji (BOM - Bills of Materials).

BIM jest opisywany jako zarządzanie informacjami o budynku, gdzie wszystkie aktualizacje i zmiany mają miejsce w bazie danych. Niezależnie od tego, czy mamy do czynienia ze schematami, przekrojami czy rysunkami arkuszy, wszystko jest zawsze skoordynowane, spójne i aktualne.

- Strona internetowa firmy CAD vendor with BIM Whitepaper, 2003 [54]

Pomysł zarządzania projektowaniem za pomocą pojedynczej zintegrowanej bazy danych był szeroko dyskutowany już w badaniach w latach 80-tych. Na przykład koncepcja BDS Charlesa Eastmana [57] zawierała 43 odniesienia do terminu "baza danych" (Rys. 6.1-2). Do 2004 r. liczba ta spadła prawie o połowę do 23 w białej księdze z 2002 r. na temat BIM [64]. W połowie XXI wieku temat baz danych praktycznie zniknął z materiałów marketingowych sprzedawców i ogólnie z agendy cyfryzacji.

Chociaż to baza danych i dostęp do niej były pierwotnie pomyślane jako rdzeń systemu BIM, z czasem nacisk przesunął się na geometrię, wizualizację i 3D. Rejestrator standardu IFC w 1994 r., który opublikował białą księgu BIM w 2002 r. - ten sam sprzedawca - w białej księdze z początku XXI wieku wyraźnie wskazał na ograniczenia neutralnych formatów, takich jak IGES, STEP i IFC oraz potrzebę bezpośredniego dostępu do baz danych CAD:

Różne aplikacje mogą być niekompatybilne, a ponownie wprowadzone dane mogą być niedokładne [...]. Rezultat tradycyjnego projektowania wspomaganego komputerowo [CAD]: wyższe koszty, dłuższy czas wprowadzenia produktu na rynek i niższa jakość produktu. Obecnie wszystkie główne aplikacje wykorzystują standardowe interfejsy branżowe do wymiany danych niskiego poziomu. Korzystając ze starych standardów IGES lub nowych STEP [IFC jest de facto i de jure kopią formatu STEP/IGES] do wymiany danych między aplikacjami różnych producentów, użytkownicy mogą osiągnąć pewną kompatybilność danych między najlepszymi w swojej klasie produktami. Jednak IGES i STEP działają tylko na niskich poziomach i nie mogą wymieniać danych tak bogatych, jak informacje generowane przez dzisiejsze wiodące aplikacje [...]. I chociaż te i inne standardy są ulepszane niemal codziennie, zawsze będą w tyle za dzisiejszymi produktami dostawców pod względem bogactwa danych. [...] Programy w ramach aplikacji muszą być w stanie wymieniać i zachowywać bogactwo danych bez uciekania się do neutralnych translatorów, takich jak IGES, STEP [IFC] lub PATRAN. Zamiast tego, aplikacje ramowe powinny mieć możliwość bezpośredniego dostępu do podstawowej bazy danych CAD, aby nie utracić szczegółowości i dokładności informacji.

- Biała księga dostawcy CAD (IFC, BIM) "Integrated Design and Manufacturing: Benefits and Rationale", 2000 [65].

Dlatego już w latach 80. i na początku XXI wieku za kluczowy element cyfrowego projektowania w środowisku CAD uznawano bazę danych, a nie plik formatu czy neutralny format IFC. Sugerowano, że należy

zrezygnować z translatorów, a aplikacje powinny mieć bezpośredni dostęp do danych. W rzeczywistości jednak w połowie lat 20. koncepcja BIM zaczęła przypominać strategię "dziel i zwyciężaj", w której interesy producentów oprogramowania korzystających z zamkniętych jąder geometrycznych są traktowane priorytetowo w stosunku do rozwoju otwartej wymiany informacji.

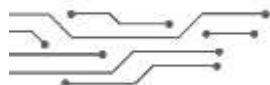
Obecnie BIM jest postrzegany jako integralna część branży budowlanej. Jednak w ciągu ostatnich dwóch dekad obietnice uproszczenia współpracy i integracji danych w dużej mierze nie zostały zrealizowane. Większość rozwiązań jest nadal powiązana z zamkniętymi lub neutralnymi formatami i specjalistycznymi narzędziami. W części 6 książki "CAD i BIM: Marketing, rzeczywistość i przyszłość danych projektowych w budownictwie" przyjrzymy się szczegółowo historii BIM, otwartego BIM i IFC, a także kwestiom interoperacyjności i jąder geometrycznych.

Obecnie branża stoi przed kluczowym wyzwaniem, jakim jest przejście od tradycyjnego rozumienia CAD (BIM) jako narzędzia do modelowania do wykorzystania go jako pełnoprawnej bazy danych. Wymaga to nowego podejścia do pracy z informacją, porzucenia zależności od zamkniętych ekosystemów i wdrożenia otwartych rozwiązań.

Wraz z rozwojem narzędzi inżynierii odwrotnej, które umożliwiają dostęp do baz danych CAD, a także rozprzestrzenianiem się technologii Open Source i LLM, użytkownicy i programiści w branży budowlanej coraz częściej odchodzą od niejasnych terminów dostawców oprogramowania. Zamiast tego uwaga skupia się na tym, co naprawdę ważne: danych (bazach danych) i procesach.

Za modnymi akronimami i wizualizacjami kryją się standardowe praktyki zarządzania danymi: przechowywanie, transfer i transformacja - czyli klasyczny proces ETL (Extract, Transform, Load). Podobnie jak w innych branżach, cyfryzacja budownictwa wymaga nie tylko standardów wymiany, ale także jasno ustrukturyzowanej obsługi heterogenicznych informacji.

Aby w pełni wykorzystać potencjał danych CAD (BIM), firmy muszą przemyśleć swoje podejście do zarządzania informacjami. Nieuchronnie doprowadzi to do kluczowego elementu cyfrowej transformacji - ujednolicenia, standaryzacji i znaczącej strukturyzacji danych, z którymi na co dzień pracują specjalisci budowlani.

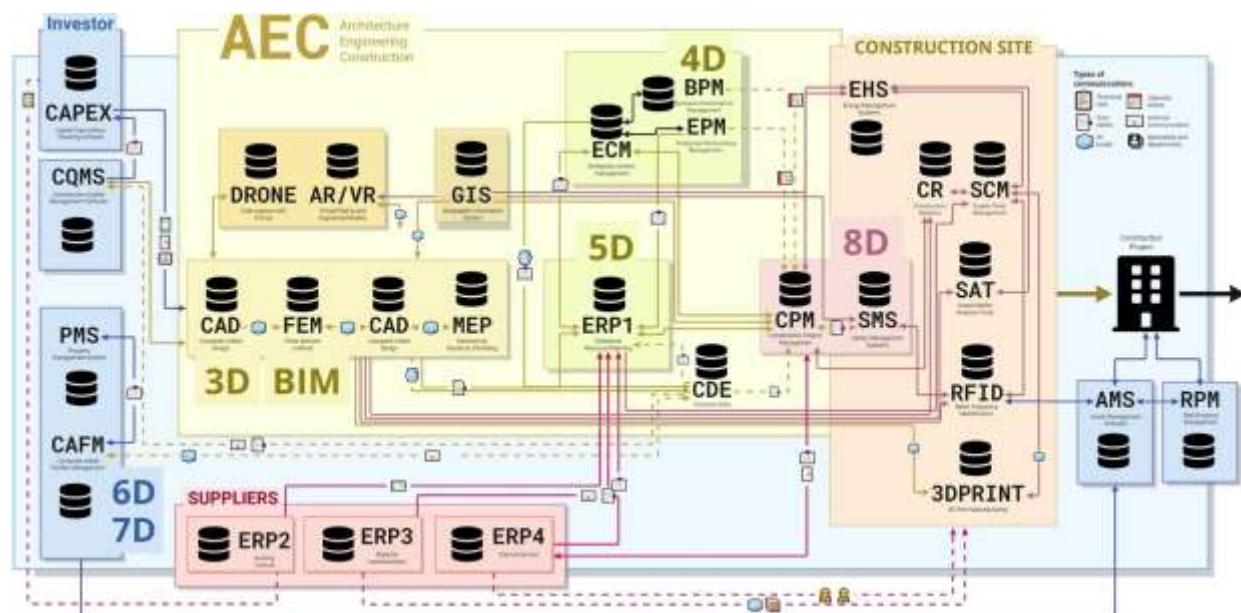


ROZDZIAŁ 2.2.

HARMONIZACJA I STRUKTURYZACJA DANYCH

Wypełnianie systemów danymi w branży budowlanej

Niezależnie od tego, czy są to duże korporacje, czy firmy średniej wielkości, specjalisci codziennie zajmują się wypełnianiem systemów oprogramowania i baz danych różnymi interfejsami z wieloformatowymi informacjami (rys. 3.2-1), które z pomocą menedżerów muszą ze sobą współpracować. To właśnie ten kompleks współdziałających systemów i procesów ostatecznie tworzy przychody i zyski dla firmy.



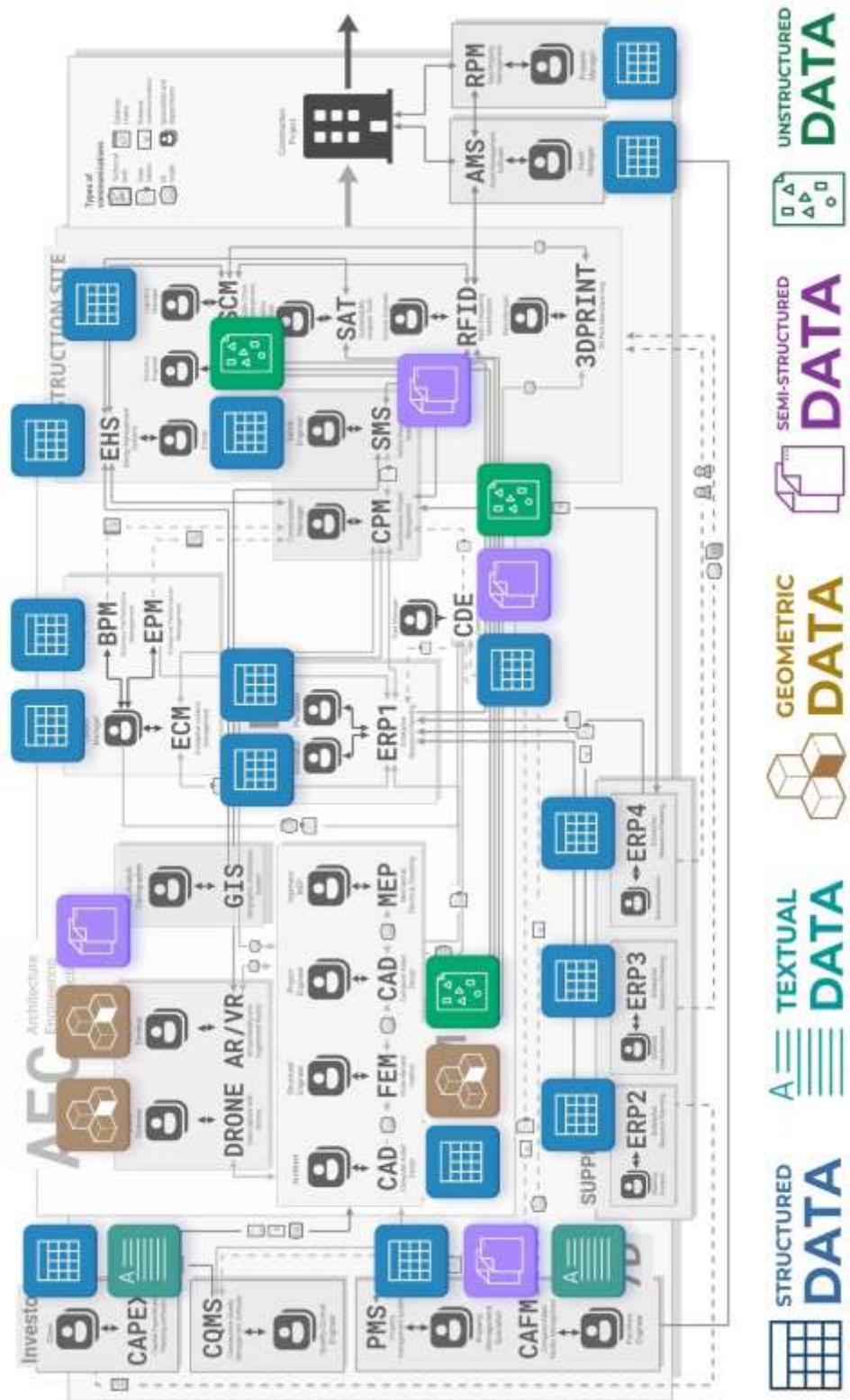
Rys. 3.2-1 Praktycznie każdy system lub aplikacja używana w branży budowlanej ma w swoim rdzeniu jedną z popularnych baz danych RDBMS.

Każda z kategorii systemów wspomnianych wcześniej i stosowanych w branży budowlanej działa z własnymi typami danych odpowiadającymi funkcjonalnej roli tych systemów. Aby przejść od poziomu abstrakcyjnego do konkretnego, przechodzimy od typów danych do ich reprezentacji w postaci formatów i dokumentów.

Do przedstawionej wcześniej listy systemów (Rys. 1.2-4) dodajemy teraz określone typy formatów i dokumentów, z którymi często pracują:

■ Inwestor (CAPEX)

- Dane finansowe: budżety, prognozy wydatków (dane ustrukturyzowane).
- Dane dotyczące trendów rynkowych: analizy rynkowe (dane ustrukturyzowane i nieustrukturyzowane).
- Dane prawne i umowne: umowy (dane tekstowe).



Rys. 3.2-2 Branża budowlana wykorzystuje wiele systemów z różnymi interfejsami, które obsługują różne typy danych.

■ **Systemy zarządzania (PMS, CAFM, CQMS)**

- Dane projektu: wykresy, zadania (dane ustrukturyzowane).
- Dane dotyczące konserwacji obiektu: plany konserwacji (dane tekstowe i częściowo ustrukturyzowane).
- Dane kontroli jakości: normy, raporty z inspekcji (dane tekstowe i nieustrukturyzowane).

■ **CAD, FEM i BIM**

- Rysunki techniczne: architektoniczne, plany konstrukcyjne (dane geometryczne, dane nieustrukturyzowane).
- Modele budynków: 3D -modele, dane materiałowe (dane geometryczne i częściowo ustrukturyzowane).
- Obliczenia inżynierijne: analiza obciążenia (dane strukturalne).

■ **Systemy zarządzania placem budowy (EHS, SCM)**

- Dane dotyczące bezpieczeństwa i zdrowia: protokoły bezpieczeństwa (dane tekstowe i ustrukturyzowane).
- Dane łańcucha dostaw: zapasy, zamówienia (dane ustrukturyzowane).
- Raporty dzienne: godziny pracy, produktywność (dane strukturalne).

■ **Drony, AR/VR, GIS, drukowanie 3D**

- Geodane: mapy topograficzne (dane geometryczne i strukturalne).
- Dane w czasie rzeczywistym: wideo i zdjęcia (dane nieustrukturyzowane).
- Modele do druku 3D: rysunki cyfrowe (dane geometryczne).

■ **Dodatkowe systemy zarządzania (4D BPM, 5D ERP1)**

- Dane dotyczące czasu i kosztów: harmonogramy, szacunki (dane ustrukturyzowane).
- Zarządzanie zmianami: zapisy zmian w projekcie (dane tekstowe i strukturalne).
- Raportowanie wyników: wskaźniki sukcesu (ustrukturyzowane dane).

■ **Integracja danych i komunikacja (CDE, RFID, AMS, RPM)**

- Wymiana danych: wymiana dokumentów, modele danych (dane strukturalne i tekstowe).
- RFID i śledzenie danych: logistyka, zarządzanie aktywami (dane ustrukturyzowane).
- Monitorowanie i kontrola: czujniki w lokalizacjach (dane ustrukturyzowane i nieustrukturyzowane).

Tak więc każdy system w branży budowlanej - od systemów zarządzania placem budowy po operacyjne bazy danych - działa z własnym typem informacji: ustrukturyzowanymi, tekstowymi, geometrycznymi i innymi. "Krajobraz danych", z którym profesjonalisci muszą pracować na co dzień, jest niezwykle zróżnicowany. Jednak proste wyliczenie formatów nie ujawnia złożoności rzeczywistej pracy z informacjami.

W praktyce firmy mają do czynienia z faktem, że dane, nawet jeśli są pobierane z systemów, nie są gotowe do użycia "tak jak są". Dotyczy to zwłaszcza tekstów, obrazów, plików PDF, plików CAD i innych formatów, które trudno analizować za pomocą standardowych narzędzi. Dlatego kolejnym kluczowym krokiem jest transformacja danych - proces, bez którego przetwarzanie, analiza, wizualizacja i podejmowanie decyzji nie mogą być skutecznie zautomatyzowane.

Transformacja danych: kluczowy fundament nowoczesnej analizy biznesowej

Obecnie większość firm staje w obliczu paradoksu: około 80% ich codziennych procesów nadal opiera się na klasycznych ustrukturyzowanych danych - znanych arkuszach kalkulacyjnych Excel i relacyjnych bazach danych (RDBMS) [66]. Jednocześnie jednak 80% nowych informacji wchodzących do cyfrowego ekosystemu firm jest nieustrukturyzowanych lub luźno ustrukturyzowanych (rys. 3.2-3) [52]. Obejmuje to tekst, grafikę, geometrię, obrazy, modele CAD, dokumentację w formacie PDF, nagrania audio i wideo, korespondencję elektroniczną i wiele innych.

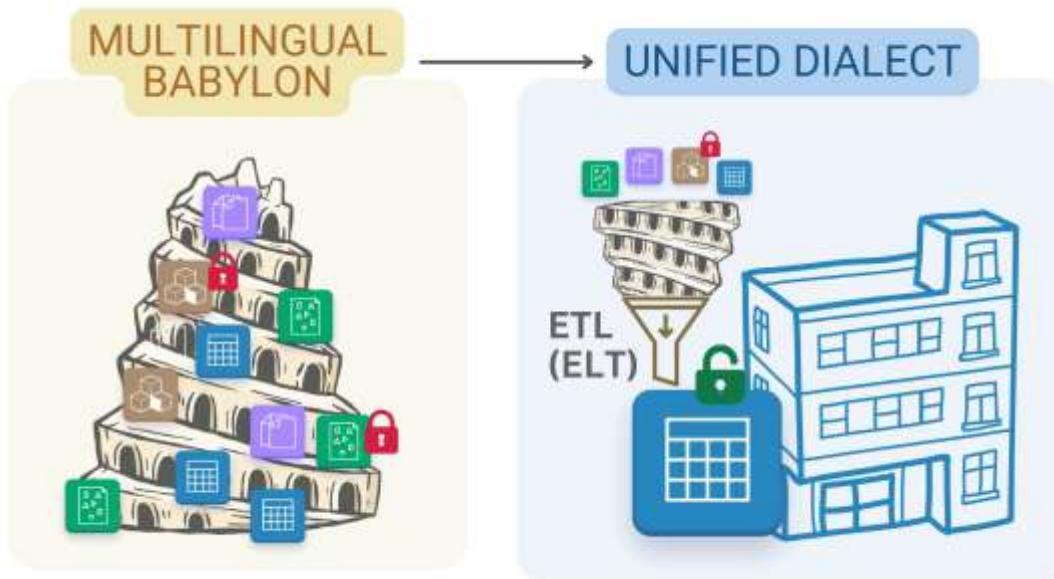
Co więcej, ilość nieustrukturyzowanych danych nadal szybko rośnie - roczna stopa wzrostu szacowana jest na 55-65% [67]. Taka dynamika stwarza poważne trudności w integracji nowych informacji z istniejącymi procesami biznesowymi. Ignorowanie tego przepływu wieloformatowych danych prowadzi do powstawania luk informacyjnych i zmniejsza możliwości zarządzania całym cyfrowym środowiskiem firmy.



Rys. 3.2-3 Coroczny wzrost ilości nieustrukturyzowanych danych stwarza wyzwania związane z integracją informacji strumieniowych z procesami biznesowymi.

Ignorowanie złożonych nieustrukturyzowanych i mylących luźno ustrukturyzowanych danych w procesach automatyzacji może prowadzić do znacznych luk w krajobrazie informacyjnym firmy. W dzisiejszym świecie niekontrolowanego i lawinowego przepływu informacji, firmy muszą przyjąć hybrydowe podejście do zarządzania danymi, które obejmuje skuteczne metody radzenia sobie ze wszystkimi typami danych.

Kluczem do skutecznego zarządzania danymi jest organizowanie, strukturyzowanie i klasyfikowanie różnych rodzajów danych "Babel" (w tym nieustrukturyzowanych, tekstowych i geometrycznych formatów, w ustrukturyzowane lub luźno ustrukturyzowane dane). Proces ten przekształca chaotyczne zbiory danych w zorganizowane struktury w celu integracji z systemami, umożliwiając tym samym podejmowanie decyzji na ich podstawie (Rys. 3.2-4).



Rys. 3.2-4 Głównym zadaniem działań zarządzania danymi jest przełożenie "Babilonu" różnorodnych i wieloformatowych danych na uporządkowany i skategoryzowany system.

Jedną z kluczowych przeszkód dla takiej harmonizacji pozostaje niski poziom interoperacyjności między różnymi platformami cyfrowymi - "silosami", o których mówiliśmy w poprzednich rozdziałach.

Zgodnie z raportem, Narodowy Instytut Standardów i Technologii (NIST, USA) podkreśla [68], że słaba kompatybilność danych pomiędzy różnymi platformami budowlanymi prowadzi do utraty informacji i znacznego dodatkowych kosztów. Tylko w 2002 r. problemy z interoperacyjnością oprogramowania spowodowały straty w amerykańskim budownictwie kapitałowym o łącznej wartości 15,8 mld USD rocznie, przy czym dwie trzecie tych strat ponoszą właściciele i operatorzy budynków, zwłaszcza podczas eksploatacji i konserwacji [68]. W badaniu zauważono również, że standaryzacja formatów danych może zmniejszyć te straty i poprawić wydajność w całym cyklu życia obiektu.

Według badania CrowdFlower z 2016 roku [69], które objęło 16 000 naukowców zajmujących się danymi na całym świecie, głównym problemem pozostają "brudne" i wieloformatowe dane. Według tego badania, najcenniejszym zasobem nie są końcowe bazy danych lub modele uczenia maszynowego, ale czas poświęcony na przygotowanie informacji.

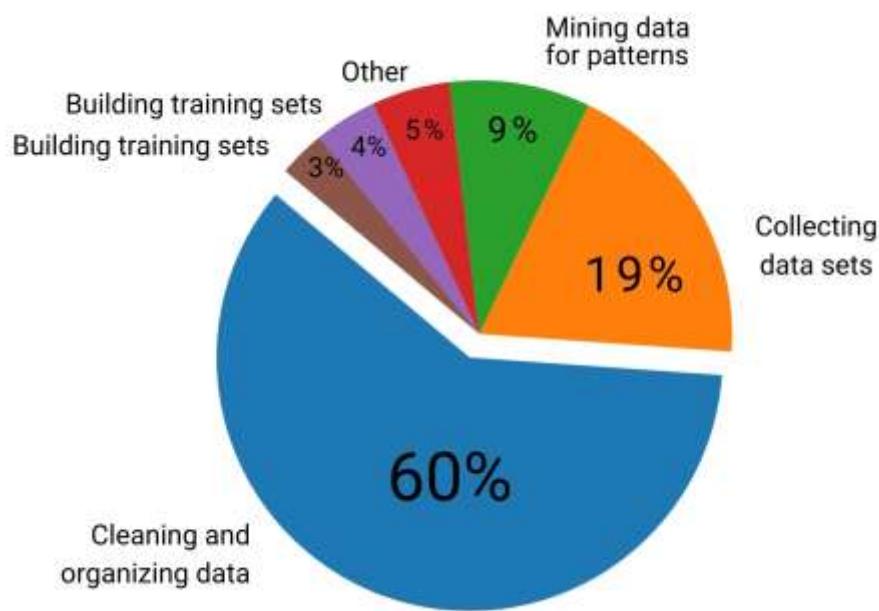
Czyszczanie, formatowanie i organizowanie danych zajmuje do 60 procent czasu analityka i menedżera danych. Prawie jedna piąta czasu poświęcana jest na wyszukiwanie i gromadzenie odpowiednich zestawów danych, które często są ukryte w silosach i niedostępne do analizy. Tylko około 9 procent czasu poświęca się bezpośrednio na modelowanie, analizę, tworzenie prognoz i testowanie hipotez. Reszta czasu poświęcana jest na komunikację, wizualizację, raportowanie i wyszukiwanie dodatkowych źródeł informacji.

Średnio praca menedżera z danymi rozkłada się w następujący sposób (Rys. 3.2-5):

- **Czyszczanie i porządkowanie danych (60%):** posiadanie czystych i ustrukturyzowanych danych może znacznie skrócić czas pracy analityka i przyspieszyć proces realizacji zadań.
- **Gromadzenie danych (19%):** Głównym wyzwaniem dla specjalistów ds. nauki o danych jest

znanie o odpowiednich zbiorach danych. Często dane firmowe są ułożone w chaotycznie zorganizowanych "silosach", co utrudnia dostęp do potrzebnych informacji.

- **Modelowanie/uczenie maszynowe (9%):** Często utrudnione przez brak jasności celów biznesowych ze strony klientów. Brak jasnej misji może zniweczyć potencjał nawet najlepszego modelu.
- **Inne zadania (5%):** oprócz przetwarzania danych analitycy muszą zajmować się badaniami, analizowaniem danych pod różnymi kątami, przekazywaniem wyników za pomocą wizualizacji i raportów oraz zalecaniem optymalizacji procesów i strategii.



Rys. 3.2-5 Na co menedżerowie danych pracujący z danymi poświęcają najwięcej czasu (na podstawie [70]).

Szacunki te znajdują potwierdzenie w innych badaniach. Według badania Xplenty opublikowanego w BizReport w 2015 roku [71], od 50% do 90% czasu specjalistów Business Intelligence (BI) (BI) poświęca się na przygotowanie danych do analizy.

Czyszczenie, walidacja i porządkowanie danych stanowi krytyczną podstawę dla wszystkich dalszych procesów związanych z danymi i analizą, zajmując do 90% czasu naukowców zajmujących się danymi.

Ta żmudna praca, niewidoczna dla użytkownika końcowego, ma kluczowe znaczenie. Błędy w surowych danych nieuchronnie znieksztalcają analizy, wprowadzają w błąd i mogą prowadzić do kosztownych błędów w zarządzaniu. Właśnie dlatego procesy czyszczenia i standaryzacji danych - od eliminacji duplikatów i uzupełniania braków po harmonizację jednostek miary i dostosowanie do wspólnego modelu - stają się kamieniem węgielnym nowoczesnej strategii cyfrowej.

Tym samym gruntowna transformacja, czyszczenie i standaryzacja danych nie tylko zajmują większość czasu specjalistów (nawet do 80% pracy z danymi), ale także determinują możliwość ich efektywnego wykorzystania w ramach nowoczesnych procesów biznesowych. Sama organizacja i czyszczenie danych nie wyczerpują jednak zadania optymalnego zarządzania przepływem informacji w firmie. Na etapie organizacji i strukturyzacji staje się wybór odpowiedniego modelu danych, który bezpośrednio wpływa na

wygodę i efektywność pracy z informacjami na kolejnych etapach przetwarzania.

Ponieważ dane i cele biznesowe są różne, ważne jest, aby zrozumieć charakterystykę modeli danych i być w stanie wybrać lub stworzyć odpowiednią strukturę. W zależności od stopnia ustrukturyzowania i sposobu opisywania relacji między elementami, istnieją trzy główne modele: ustrukturyzowany, luźno ustrukturyzowany i graficzny. Każdy z nich nadaje się do innych zadań i ma swoje mocne i słabe strony.

Modele danych: relacje w danych i relacje między elementami

Dane w systemach informatycznych są organizowane na różne sposoby - w zależności od zadań i wymagań dotyczących przechowywania, przetwarzania i przesyłania informacji. Kluczową różnicą między typami modeli danych, formą, w jakiej przechowywane są informacje, jest stopień ustrukturyzowania i sposób, w jaki opisywane są relacje między elementami.

Dane strukturalne mają jasny i powtarzalny schemat: są zorganizowane jako tabele z ustalonymi kolumnami. Format ten zapewnia przewidywalność, łatwość przetwarzania i wydajność podczas wykonywania zapytań SQL, filtrowania i agregacji. Przykłady - bazy danych (RDBMS), Excel, CSV.

Luźno ustrukturyzowane dane umożliwiają elastyczną strukturę: różne elementy mogą zawierać różne zestawy atrybutów i być przechowywane jako hierarchie. Przykładami są JSON, XML lub inne formaty dokumentów. Dane te są wygodne, gdy konieczne jest modelowanie zagnieżdżonych obiektów i relacji między nimi, ale z drugiej strony komplikują analizę i standaryzację danych (rys. 3.2-6).

	Data Model	Storage Format	Example
	Relational	CSV, SQL	A table of doors in Excel
	Hierarchical	JSON, XML	Nested door objects inside a room
	Graph-based	RDF, GraphDB	Relationships between building elements

Rys. 3.2-6 Model danych to logiczna struktura opisująca sposób organizacji, przechowywania i przetwarzania danych w systemie.

Wybór odpowiedniego formatu zależy od celów:

- Jeśli ważna jest szybkość filtrowania i analizy - wystarczą tabele relacyjne (SQL, CSV, RDBMS, kolumnowe bazy danych).
- Jeśli wymagana jest elastyczność struktury - lepiej jest użyć JSON lub XML.
- Jeśli dane mają złożone relacje - grafowe bazy danych zapewniają widoczność i skalowalność.

W klasycznych relacyjnych bazach danych (RDBMS) każda encja (np. drzwi) jest reprezentowana przez wiersz, a jej właściwości przez kolumny tabeli. Na przykład tabela elementów z kategorii "Drzwi" może zawierać pola ID, Wysokość, Szerokość, Odporność ogniowa i ID pomieszczenia wskazujące pomieszczenie (rys. 3.2-7).

W klasycznych relacyjnych bazach danych (RDBMS) relacje są tworzone w postaci tabel, gdzie każdy rekord reprezentuje obiekt, a kolumny reprezentują jego parametry. W formacie tabelarycznym dane o drzwiach w projekcie wyglądają następująco, gdzie każdy wiersz reprezentuje oddzielny element - drzwi z unikalnym identyfikatorem i atrybutami, a połączenie z pokojem odbywa się za pomocą parametru "Room ID".



Door ID	Room ID	Height (mm)	Width (mm)	Fireproof
ID1001	101	2000	900	Yes
ID1002	101	2100	800	No
ID1003	102	2000	850	Yes

Rys. 3.2-7 Informacje o trzech elementach kategorii "Drzwi" projektu w formie tabelarycznej.

W luźno ustrukturyzowanych formatach, takich jak JSON lub XML, dane są przechowywane w formie hierarchicznej lub zagnieżdżonej, gdzie elementy mogą zawierać inne obiekty, a ich struktura może się różnić. Pozwala to na modelowanie złożonych relacji między elementami. Podobne informacje o drzwiach w projekcie, które zostały zapisane w ustrukturyzowanej formie (Rys. 3.2-7), są reprezentowane w luźno ustrukturyzowanym formacie (JSON) w taki sposób (Rys. 3.2-8), że stają się zagnieżdzonymi obiektami w pokojach (Pokoje - ID), co logicznie odzwierciedla hierarchię.

```

1 {
2   "Rooms": [
3     {
4       "ID": 101,
5       "Doors": [
6         {"ID": 1, "Height": 2000, "Width": 900, "Fireproof": "Yes"},
7         {"ID": 2, "Height": 2100, "Width": 800, "Fireproof": "No"}
8       ]
9     },
10    {
11      "ID": 102,
12      "Doors": [
13        {"ID": 3, "Height": 2000, "Width": 850, "Fireproof": "Yes"}
14      ]
15    }
16  ]
17 }
18

```

Rys. 3.2-8 Informacje o elementach kategorii "Drzwi" projektu w formacie JSON.

W modelu grafu dane są reprezentowane jako węzły (wierzchołki) i połączenia (krawędzie) między nimi. Pozwala to na wizualizację złożonych relacji między obiektami i ich atrybutami. W przypadku danych drzwi i pomieszczeń w projekcie, reprezentacja grafu wygląda następująco:

- **Węzły (nodes)** reprezentują główne jednostki: pokoje (Room 101, Room 102) i drzwi (ID1001, ID1002, ID1003)
- **Żebra (linki)** pokazują relacje między tymi jednostkami, np. przynależność drzwi do określonego pomieszczenia.
- **Atrybuty** są mapowane na węzły i zawierają właściwości jednostki (wysokość, szerokość, odporność ogniowa drzwi).



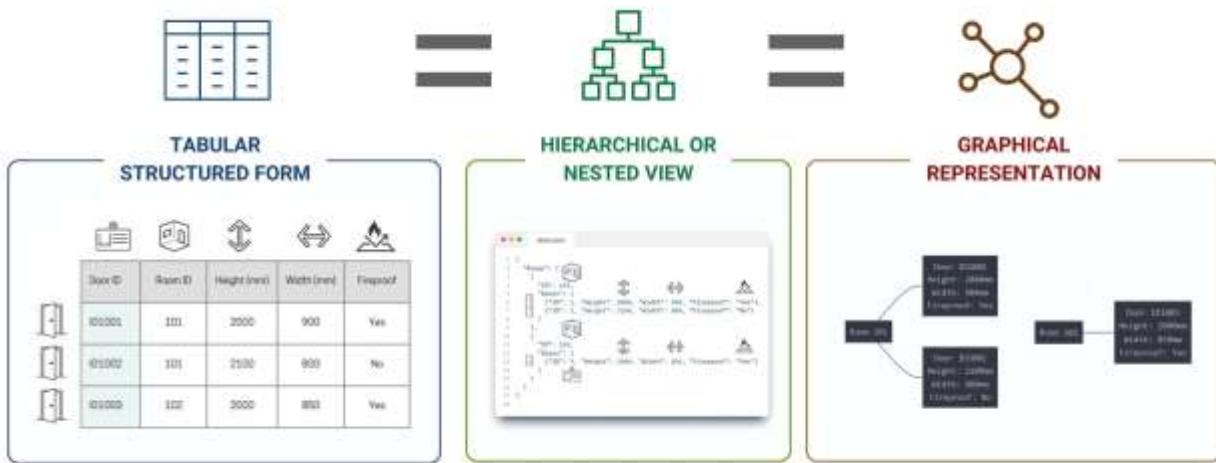
Rys. 3.2-9 Informacje o jednostce Project door w widoku wykresu.

W grafowym modelu danych opisu drzwi, każde pomieszczenie i każde drzwi są oddzielnymi węzłami. Drzwi

są połączone z pomieszczeniami za pomocą krawędzi, które wskazują, że drzwi należą do określonego pomieszczenia. Atrybuty drzwi (wysokość, szerokość, odporność ogniodziałająca) są przechowywane jako właściwości odpowiednich węzłów. Więcej szczegółów na temat formatów grafów i tego, jak semantyka grafów pojawiła się w branży budowlanej, zostanie omówionych w rozdziale "Pojawienie się semantyki i ontologii w budownictwie".

Grafowe bazy danych są skuteczne, gdy ważne są nie tyle same dane, co relacje między nimi, np. w systemach rekomendacji, systemach routingu lub podczas modelowania złożonych relacji w projektach zarządzania obiektami. Format grafu upraszcza tworzenie nowych relacji, umożliwiając dodawanie nowych typów danych do grafu bez zmiany struktury przechowywania. Jednak w porównaniu do tabel relacyjnych i formatów strukturalnych, w grafie nie ma dodatkowej łączności danych - przeniesienie dwuwymiarowych danych bazy danych do grafu nie zwiększa liczby relacji i nie pozwala na uzyskanie nowych informacji.

Forma i schemat danych powinny być dostosowane do konkretnego przypadku użycia i zadań do wykonania. Aby efektywnie pracować w procesach biznesowych, ważne jest, aby korzystać z tych narzędzi i tych modeli danych, które pomagają uzyskać wyniki tak szybko i łatwo, jak to możliwe.



Rys. 3.2-10 Te same informacje o elementach projektu mogą być przechowywane w różnych formatach przy użyciu różnych modeli danych.

Obecnie większość dużych firm boryka się z problemem nadmiernej złożoności danych. Każda z setek lub tysięcy aplikacji korzysta z własnego modelu danych, co powoduje nadmierną złożoność - pojedynczy model jest często dziesiątki razy bardziej złożony niż to konieczne, a agregat wszystkich modeli jest tysiące razy bardziej złożony. Ta nadmierna złożoność znacznie utrudnia pracę zarówno programistom, jak i użytkownikom końcowym.

Taka złożoność nakłada poważne ograniczenia na rozwój i utrzymanie systemów firmy. Każdy nowy element w modelu wymaga dodatkowego kodu, wdrożenia nowej logiki, dokładnego przetestowania i dostosowania do istniejących rozwiązań. Wszystko to zwiększa koszty i spowalnia pracę zespołu automatyzacji w firmie, zamieniając nawet proste zadania w kosztowne i czasochłonne procesy.

Złożoność wpływa na wszystkie poziomy architektury danych. W relacyjnych bazach danych wyraża się w rosnącej liczbie tabel i kolumn, często nadmiarowych. W systemach obiektowych złożoność jest zwiększała przez mnogość klas i wzajemnie powiązanych właściwości. W formatach takich jak XML czy JSON złożoność przejawia się w zagmatwanych, zagnieżdżonych strukturach, unikalnych kluczach i niespójnych schematach.

Nadmierna złożoność modeli danych sprawia, że systemy są nie tylko mniej wydajne, ale także trudne do zrozumienia przez użytkowników końcowych, a w przyszłości duże modele językowe i agenty LLM. To właśnie problem zrozumienia i złożoności modeli danych i przetwarzania danych rodzi pytanie: jak sprawić, by dane były na tyle łatwe w użyciu, by szybko zaczęły być użyteczne.

Nawet jeśli modele danych są wybierane mądrze, ich użyteczność jest znacznie ograniczona, jeśli dostęp do danych jest ograniczony. Zastrzeżone formaty i zamknięte platformy utrudniają integrację, komplikują automatyzację i odbierają kontrolę nad zastrzeżonymi informacjami, tworząc nie tylko silos nowych danych, ale także zamknięty silos, do którego dostęp można uzyskać tylko po uzyskaniu autoryzacji od dostawcy. Aby zrozumieć skalę problemu, ważne jest, aby dokładnie rozważyć, w jaki sposób zamknięte systemy wpływają na procesy cyfrowe w budownictwie.

Zastrzeżone formaty i ich wpływ na procesy cyfrowe

Jednym z kluczowych wyzwań stojących przed firmami budowlanymi podczas cyfryzacji jest ograniczony dostęp do danych. Utrudnia to integrację systemów, obniża jakość informacji i utrudnia organizację wydajnych procesów. U podstaw tych trudności często leży stosowanie zastrzeżonych formatów i zamkniętych rozwiązań programowych.

Niestety, do tej pory wiele programów wykorzystywanych w branży budowlanej pozwala użytkownikowi jedynie na zapisywanie danych w zastrzeżonych formatach lub w chmurze, do których dostęp można uzyskać jedynie za pośrednictwem ściśle ograniczonych interfejsów. Nierzadko rozwiązania te budowane są w oparciu o jeszcze bardziej zamknięte systemy pochodzące od większych dostawców. W rezultacie nawet ci deweloperzy, którzy chcieliby oferować bardziej otwarte architektury, są zmuszeni do przestrzegania zasad dyktowanych przez dużych dostawców.

Podczas gdy nowoczesne systemy zarządzania danymi budowlanymi w coraz większym stopniu obsługują otwarte formaty i standardy (rys. 3.1-5), bazy danych oparte na CAD (BIM) oraz powiązane systemy ERP i CAFM pozostają odizolowanymi, zastrzeżonymi "wyspami" w cyfrowym krajobrazie branży (rys. 3.2-11).

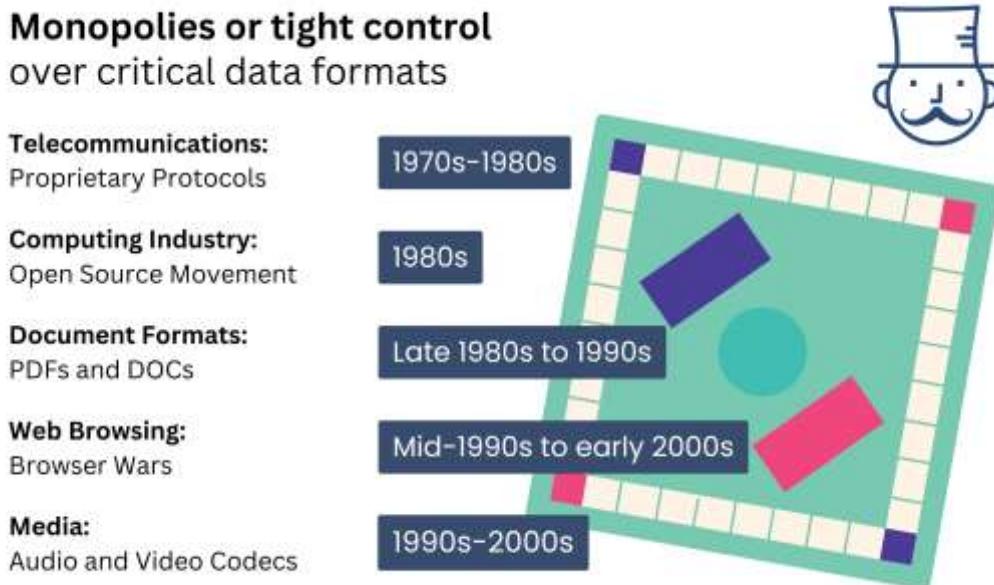


Rys. 3.2-11 Zamknięty i zastrzeżony charakter danych stwarza bariery dla integracji i dostępu do danych.

Zamknięte i zmonopolizowane formaty i protokoły to nie tylko problem branży budowlanej. W wielu sektorach gospodarki walka z zamkniętymi standardami i ograniczonym dostępem do danych rozpoczęła się od spowolnienia innowacji (Rys. 3.2-12), istnienia sztucznych barier wejścia dla nowych graczy i pogłębiającej się zależności od dużych dostawców. Wraz z szybkim wzrostem znaczenia danych, organy ochrony konkurencji po prostu nie mają czasu na reagowanie na wyzwania stawiane przez nowe rynki cyfrowe, w wyniku czego zamknięte formaty i ograniczony dostęp do danych zasadniczo stają się cyfrowymi "granicami", które ograniczają przepływ informacji i wzrost [63].

Jeśli maszyny będą produkować wszystko, czego potrzebujemy, nasza sytuacja będzie zależeć od tego, jak te dobra będą dystrybuowane. Każdy będzie mógł cieszyć się dostatnim życiem tylko wtedy, gdy bogactwo wytwarzane przez maszyny będzie dzielone. Albo większość ludzi skończy żyjąc w skrajnym ubóstwie, jeśli właściciele samochodów będą skutecznie lobbywać przeciwko redystrybucji bogactwa. Jak dotąd wydaje się, że sprawy idą w tym drugim kierunku, a technologia prowadzi do coraz większych nierówności [72].

- Stephen Hawking, astrofizyk, 2015 r.



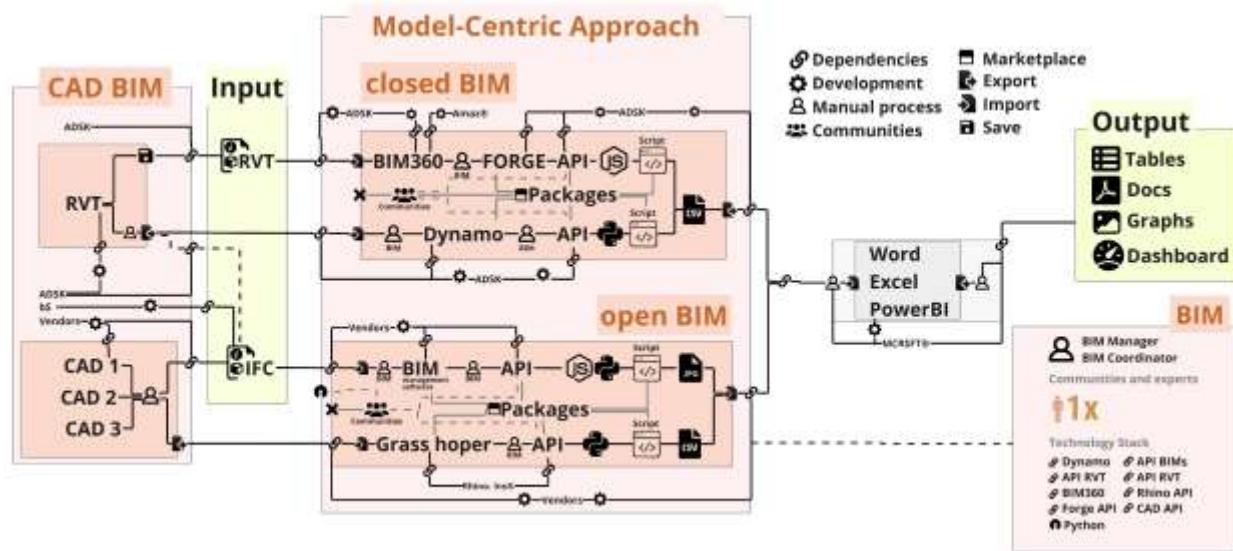
Rys. 3.2-12 Monopolistyczna własność kluczowych formatów danych i protokołów nie jest wyłącznym problemem branży budowlanej.

W rezultacie, ze względu na zamknięty dostęp do baz danych programów, menedżerowie danych, analitycy danych, specjalisci IT i programiści tworzący aplikacje do dostępu do danych, przetwarzania i automatyzacji w branży budowlanej stoją dziś w obliczu licznych zależności od dostawców oprogramowania (rys. 3.2-13). Zależności te w postaci dodatkowych warstw dostępu wymagają tworzenia rozwiązań z wyspecjalizowanymi API -połączeniami oraz specjalnymi narzędziami i oprogramowaniem.

API (Application Programming Interface) to sformalizowany interfejs, za pośrednictwem którego jeden program może wchodzić w interakcję z innym, wymieniając dane i funkcje bez konieczności uzyskiwania dostępu do kodu źródłowego. API opisuje, jakie żądania może wykonać zewnętrzny system, w jakim formacie powinny one być i jakie odpowiedzi otrzyma. Jest to znormalizowana "umowa" między modułami oprogramowania.

Duża liczba zależności od zamkniętych rozwiązań powoduje, że cała architektura kodu i logika procesów biznesowych w firmie staje się "architekturą spaghetti" narzędzi zależnych od polityki dostawcy oprogramowania w zakresie zapewnienia wysokiej jakości dostępu do danych.

Uzależnienie od zamkniętych rozwiązań i platform prowadzi nie tylko do utraty elastyczności, ale także do realnego ryzyka biznesowego. Zmiana warunków licencyjnych, zamknięcie dostępu do danych, zmiana formatów lub struktury API - wszystko to może zablokować krytyczne procesy. Nagle okazuje się, że aktualizacja jednej tabeli wymaga przerobienia całego bloku integracji i konektorów (rys. 3.2-13), a każda duża aktualizacja oprogramowania lub jego dostawcy API staje się potencjalnym zagrożeniem dla stabilności całego systemu firmy.



Rys. 3.2-13 Przykład dużej liczby zależności w przetwarzaniu CAD -danych tworzy bariery dla integracji danych w ekosystemie firm budowlanych.

Programiści i architekci systemów w takich warunkach są zmuszeni pracować nie dla przewidywania, ale dla przetrwania. Zamiast wdrażać nowe rozwiązania, dostosowują się. Zamiast rozwijać, starają się zachować kompatybilność. Zamiast automatyzować i przyspieszać procesy, spędzają czas na studiowaniu kolejnych zamkniętych interfejsów, dokumentacji API i niekończącej się przebudowie kodu.

Praca z zamkniętymi formatami i systemami to nie tylko wyzwanie techniczne - to strategiczne ograniczenie. Pomimo oczywistych możliwości oferowanych przez nowoczesną automatyzację, sztuczną inteligencję, LLM i analitykę predykcyjną, wiele firm nie wykorzystuje w pełni swojego potencjału. A bariery wzniesione przez zastrzeżone formaty (Rys. 3.2-13) uniemożliwiają firmom dostęp do ich własnych danych. Jest to być może ironia cyfrowej transformacji w budownictwie.

Przejrzystość danych i otwarte systemy nie są luksusem, ale warunkiem wstępny szybkości i wydajności. Bez otwartości procesy biznesowe są wypełnione niepotrzebną biurokracją, wielowarstwowymi łańcuchami zatwierdzania i rosnącą zależnością od zasady HiPPO - podejmowania decyzji w oparciu o opinię najwyższej opłacanej osoby.

Niemniej jednak na horyzoncie rysuje się zmiana paradygmatu. Pomimo dominacji rozwiązań własnościowych, coraz więcej firm zdaje sobie sprawę z ograniczeń architektur inspirowanych czwartą rewolucją przemysłową. Obecnie wektor przesuwa się w kierunku zasad piątej rewolucji, w której dane jako zasób strategiczny, otwarte interfejsy (API) i prawdziwa interoperacyjność między systemami znajdują się w centrum uwagi.

Przejście to oznacza odejście od zamkniętych ekosystemów na rzecz elastycznych, modułowych architektur cyfrowych, w których kluczowe znaczenie mają otwarte formaty, standardy i przejrzysta wymiana danych.

Otwarte formaty zmieniają podejście do cyfryzacji

Branża budowlana była jedną z ostatnich, które zajęły się problemem zamkniętych i zastrzeżonych danych. W przeciwieństwie do innych sektorów gospodarki, cyfryzacja rozwijała się tu powoli. Powodem tego jest tradycyjny, konserwatywny charakter branży, powszechność rozproszonych rozwiązań lokalnych oraz głęboko zakorzeniony charakter zarządzania dokumentami w formie papierowej. Przez dziesięciolecia kluczowe procesy budowlane opierały się na fizycznych rysunkach, rozmowach telefonicznych i niezsynchonizowanych bazach danych. W tym kontekście zamknięte formaty przez długi czas były postrzegane jako norma, a nie przeszkoda.

Doświadczenia z innych branż pokazują, że usunięcie barier dla zamkniętych danych prowadzi do wzrostu innowacyjności, przyspieszenia rozwoju i zwiększenia konkurencji [73]. W nauce wymiana otwartych danych pozwala na przyspieszenie odkryć i rozwój współpracy międzynarodowej. W medycynie może poprawić skuteczność diagnostyki i leczenia. W inżynierii oprogramowania - do tworzenia ekosystemów współtworzenia i szybkiego ulepszania produktów.

Według raportu McKinsey "Open Data: Unlock Innovation and Productivity with Information Flow" z 2013 r. [74], otwarte dane mają potencjał odblokowania od 3 do 5 bilionów dolarów rocznie w siedmiu kluczowych branżach, w tym w budownictwie, transporcie, opiece zdrowotnej i energetyce. Według tego samego badania, zdecentralizowane ekosystemy danych umożliwiają dużym firmom budowlanym i wykonawcom zmniejszenie kosztów rozwoju i utrzymania oprogramowania, przyspieszając wdrażanie technologii cyfrowych.

Przejście na otwarte architektury, które już dawno rozpoczęło się w innych sektorach gospodarki, stopniowo rozprzestrzenia się na branżę budowlaną. Duże firmy i klienci publiczni, a zwłaszcza organizacje finansowe, które kontrolują inwestycje w projekty budowlane, coraz częściej domagają się wykorzystania otwartych danych i dostępu do kodu źródłowego obliczeń, kalkulacji i aplikacji. Od deweloperów nie oczekuje się już tylko tworzenia rozwiązań cyfrowych i pokazywania ostatecznych danych liczbowych projektu - oczekuje się od nich przejrzystości, powtarzalności i niezależności od zewnętrznych dostawców aplikacji.

Korzystanie z rozwiązań open source daje klientowi pewność, że nawet jeśli zewnętrzni deweloperzy przestaną współpracować lub opuszczą projekt, nie wpłynie to na możliwość dalszego rozwoju narzędzi i systemów. Jedną z głównych zalet otwartych danych jest możliwość wyeliminowania zależności twórców aplikacji od konkretnych platform dostępu do danych.

Jeśli firma nie może całkowicie zrezygnować z zastrzeżonych rozwiązań, możliwym kompromisem jest zastosowanie technik inżynierii odwrotnej. Te legalne i technicznie uzasadnione metody pozwalają na przekształcenie zamkniętych formatów w bardziej dostępne, ustrukturyzowane i nadające się do integracji. Jest to szczególnie ważne, gdy istnieje potrzeba połączenia ze starszymi systemami lub migracji informacji z jednego środowiska oprogramowania do innego.

Jednym z najjaśniejszych przykładów w historii przechodzenia na otwarte formaty i stosowania inżynierii odwrotnej (legalnego hakowania zastrzeżonych systemów) w budownictwie jest historia walki o otwarcie formatu DWG, szeroko stosowanego w systemach projektowania wspomaganego komputerowo (CAD). W 1998 r., w odpowiedzi na monopol jednego producenta oprogramowania, pozostałych 15 producentów CAD utworzyło nowy sojusz o nazwie "Open DWG", aby zapewnić programistom bezpłatne i niezależne narzędzia do pracy z formatem DWG (de facto standardem przesyłania rysunków) bez potrzeby korzystania z zastrzeżonego oprogramowania lub zamkniętych interfejsów API. Wydarzenie to było punktem zwrotnym, który pozwolił dziesiątkom tysięcy firm uzyskać bezpłatny dostęp do zamkniętego formatu popularnego rozwiązań CAD od końca lat 80. do dziś i stworzyć kompatybilne rozwiązania, które sprzyjały konkurencji na rynku CAD [75]. Obecnie "Open DWG" SDK, który został stworzony w 1996 roku, jest używany w prawie wszystkich rozwiązaniach, w których można importować, edytować i eksportować format DWG, poza oficjalną aplikacją twórcy formatu DWG.

Inni giganci technologiczni wymuszają podobne transformacje. Microsoft, niegdyś symbol zastrzeżonego podejścia, otworzył kod źródłowy .NET Framework, zaczął używać Linuksa w infrastrukturze usług w chmurze Azure i przejął GitHub, aby wzmacnić swoją pozycję w społeczności Open Source. [76]. Meta (dawniej Facebook) opublikowała modele AI typu open source, takie jak seria Llama, w celu wspierania innowacji i współpracy w zakresie rozwoju agentów AI. Dyrektor generalny Mark Zuckerberg przewiduje, że platformy open source będą wiodły prym w postępie technologicznym w ciągu następnej dekady [77].

Open Source to model rozwoju i dystrybucji oprogramowania, w którym kod źródłowy jest otwarty do swobodnego użytku, badania, modyfikacji i dystrybucji.

Otwarte dane i rozwiązania open source stają się nie tylko trendem, ale podstawą zrównoważonego rozwoju cyfrowego. Zapewniają one firmom elastyczność, odporność, kontrolę nad własnymi decyzjami i możliwość skalowania procesów cyfrowych bez uzależnienia od polityki dostawców. I, co równie ważne, przywracają firmom kontrolę nad najcenniejszym zasobem XXI wieku - ich danymi.

Zmiana paradygmatu: Open Source jako koniec ery dominacji producentów oprogramowania

Branża budowlana przechodzi zmianę, której nie da się spieniżyć w zwykły sposób. Koncepcja oparta na danych, podejście skoncentrowane na danych i wykorzystanie narzędzi Open Source prowadzi do ponownego przemyślenia zasad gry, na których stoją giganci oprogramowania na rynku.

W przeciwieństwie do poprzednich transformacji technologicznych, przejście to nie będzie aktywnie promowane przez sprzedawców. Zmiana paradygmatu zagraża ich tradycyjnym modelom biznesowym opartym na licencjach, subskrypcjach i doradztwie. Nowa rzeczywistość nie obejmuje gotowego produktu ani płatnej subskrypcji - wymaga dostosowania procesów i sposobu myślenia.

Aby zarządzać i rozwijać rozwiązania centrów danych oparte na otwartych technologiach, firmy będą musiały przemyśleć wewnętrzne procesy. Specjalisci z różnych działów będą musieli nie tylko

współpracować, ale także przemyśleć sposób, w jaki ze sobą współpracują.

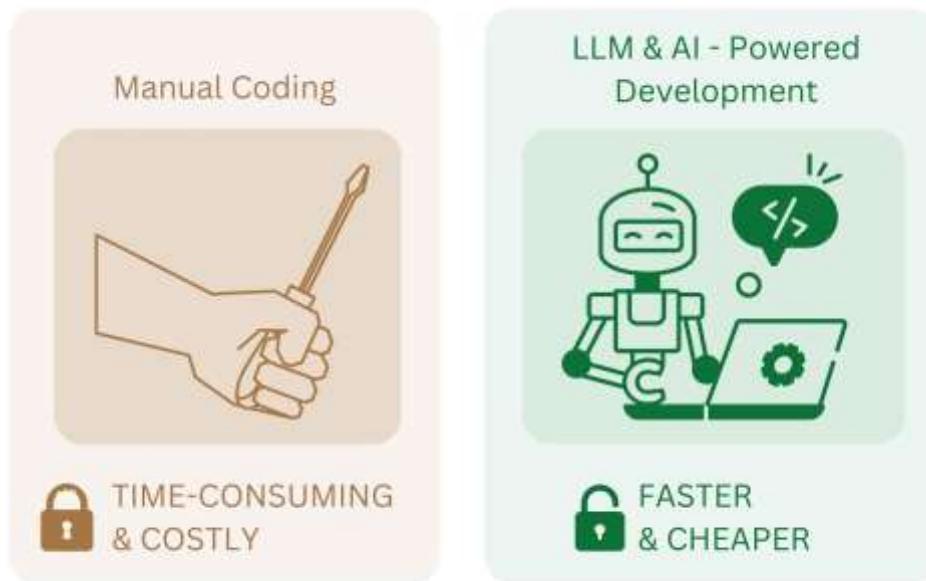
Nowy paradymat zakłada wykorzystanie otwartych danych i rozwiązań Open Source, gdzie szczególną rolę w tworzeniu kodu oprogramowania będą odgrywać narzędzia oparte na sztucznej inteligencji i dużych modelach językowych (LLM), a nie programiści. Już w połowie 2024 roku ponad 25 procent nowego kodu w Google będzie tworzony przy użyciu sztucznej inteligencji [78]. W przyszłości kodowanie za pomocą LLM będzie wykonywać 80% pracy w zaledwie 20% czasu (Rys. 3.2-14).

Według badania McKinsey's 2020 [79], układy GPU coraz częściej zastępują procesory centralne w analityce ze względu na ich wysoką wydajność i wsparcie ze strony nowoczesnych narzędzi Open Source. Pozwala to firmom przyspieszyć przetwarzanie danych bez znaczących inwestycji w drogie oprogramowanie lub zatrudniania rzadkich specjalistów.

Wiodące firmy konsultingowe, takie jak McKinsey, PwC i Deloitte, podkreślają rosnące znaczenie otwartych standardów, aplikacji Open Source w różnych branżach.

Według raportu PwC Open Source Monitor 2019 [80], 69% firm zatrudniających 100 lub więcej pracowników świadomie korzysta z rozwiązań Open Source. OSS jest szczególnie aktywnie wykorzystywany w dużych firmach: 71% firm zatrudniających 200-499 pracowników, 78% w kategorii 500-1999 pracowników i aż 86% wśród firm zatrudniających ponad 2000 pracowników. Według raportu Synopsys OSSRA 2023, 96% analizowanych baz kodu zawierało komponenty open source [81].

W przyszłości rolą programisty nie będzie ręczne pisanie kodu, ale projektowanie modeli danych, architektur przepływu i zarządzanie agentami AI, które tworzą odpowiednie obliczenia na żądanie. Interfejsy użytkownika staną się minimalistyczne, a interakcja będzie oparta na dialogu. Klasyczne programowanie ustąpi miejsca projektowaniu wysokiego poziomu i orkiestracji rozwiązań cyfrowych (Rys. 3.2-14). Obecne trendy - takie jak platformy niskokodowe (rys. 7.4-6) i ekosystemy oparte na LLM (rys. 7.4-4) - znacznie obniżają koszty rozwoju i utrzymania systemów informatycznych.



Rys. 3.2-14 Podczas gdy obecnie aplikacje są tworzone ręcznie przez programistów, w przyszłości znaczna część kodu będzie generowana przez sztuczną inteligencję i rozwiązania oparte na LLM.

Ta transformacja będzie inna niż poprzednie, a wielcy producenci oprogramowania raczej nie będą katalizatorami.

W badaniu Harvard Business School "The Value of Open Source Software" 2024 [40], całkowita wartość oprogramowania open source jest szacowana z dwóch punktów widzenia. Z jednej strony, jeśli obliczymy, ile zajęłoby zbudowanie od podstaw wszystkich istniejących rozwiązań Open Source, kwota ta wyniosłaby około 4,15 miliarda dolarów. Z drugiej strony, jeśli wyobrażymy sobie, że każda firma samodzielnie opracowuje własne analogi rozwiązań Open Source (co zdarza się wszędzie), bez dostępu do istniejących narzędzi, wówczas całkowity koszt działalności osiągnąłby kolosalną kwotę 8,8 biliona dolarów - jest to koszt popytu.

Nietrudno zgadnąć, że żaden z głównych dostawców oprogramowania nie jest zainteresowany zmniejszeniem rynku oprogramowania o potencjalnej wartości 8,8 biliona dolarów do zaledwie 4,15 miliarda dolarów. Oznaczałoby to zmniejszenie wielkości popytu ponad 2000 razy. Taka transformacja jest po prostu nieopłacalna dla dostawców, których modele biznesowe opierają się na wieloletnim utrzymywaniu zależności klientów od zamkniętych rozwiązań. Tak więc firmy oczekujące, że ktoś zaoferuje im wygodne i otwarte rozwiązanie "pod klucz", mogą być rozczarowane - tacy dostawcy po prostu się nie pojawią.

Przejście na otwartą architekturę cyfrową nie oznacza utraty miejsc pracy ani przychodów. Wręcz przeciwnie, stwarza warunki dla elastycznych i adaptacyjnych modeli biznesowych, które mogą ostatecznie wyprzeć tradycyjny rynek licencji i oprogramowania pudełkowego.

Zamiast sprzedaży licencji - usługi, zamiast zamkniętych formatów - otwarte platformy, zamiast uzależnienia od dostawcy - niezależność i możliwość budowania rozwiązań pod rzeczywiste potrzeby. Ci, którzy do tej pory po prostu korzystali z narzędzi, będą mogli stać się ich współautorami. A ci, którzy

potrafią pracować z danymi, modelami, scenariuszami i logiką, znajdą się w centrum nowej cyfrowej gospodarki branży. Więcej o tych zmianach i o tym, jakie nowe role, modele biznesowe i formaty współpracy pojawiają się wokół otwartych danych, opowiem w ostatniej, dziesiątej części książki.

Rozwiązania oparte na otwartych danych i otwartym kodzie pozwolą firmom skupić się na wydajności procesów biznesowych, zamiast na zmaganiu się z przestarzałymi interfejsami API i integracją zamkniętych systemów. Świadome przejście na otwartą architekturę może znaczco poprawić produktywność i zmniejszyć zależność od dostawców.

Przejście do nowej rzeczywistości to nie tylko zmiana podejścia do tworzenia oprogramowania, ale także przemyślenie samej zasady pracy z danymi. W centrum tej transformacji nie znajduje się kod, ale informacja: jej struktura, dostępność i możliwość interpretacji. W tym miejscu otwarte i ustrukturyzowane dane wysuwają się na pierwszy plan, stając się integralną częścią nowej architektury cyfrowej.

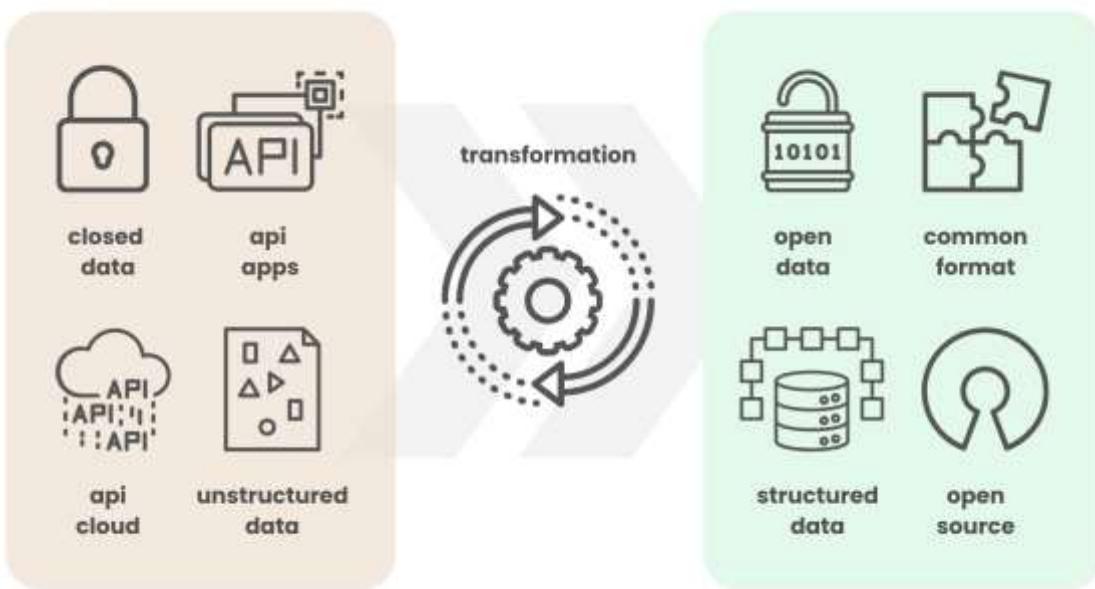
Ustrukturyzowane otwarte dane: fundament cyfrowej transformacji

Podczas gdy w poprzednich dekadach zrównoważony rozwój biznesu był w dużej mierze zdeterminowany przez wybór rozwiązań programowych i zależność od konkretnych dostawców, w dzisiejszej gospodarce cyfrowej kluczowym czynnikiem jest jakość danych i zdolność do efektywnej pracy z nimi. Otwarty kod źródłowy jest ważną częścią nowego paradymatu technologicznego, ale jego potencjał jest naprawdę odblokowany tylko wtedy, gdy dane są zrozumiałe, zorganizowane i nadają się do odczytu maszynowego. Spośród wszystkich rodzajów modeli danych, ustrukturyzowane otwarte dane stają się kamieniem węgielnym zrównoważonej transformacji cyfrowej.

Główna zaleta ustrukturyzowanych otwartych danych jest jednoznaczna interpretacja i możliwość zautomatyzowanego przetwarzania. Umożliwia to znaczny wzrost wydajności zarówno na poziomie poszczególnych operacji, jak i całej organizacji.

Według raportu Deloitte "Proces transferu danych w transformacji przedsiębiorstwa" [82], współpraca z działem IT w celu zarządzania transferem ustrukturyzowanych danych ma kluczowe znaczenie. Według brytyjskiego raportu rządowego "Data Analytics and AI in Government Project Delivery" (2024) [83], usunięcie barier w udostępnianiu danych między różnymi projektami i organizacjami ma kluczowe znaczenie dla poprawy efektywności zarządzania projektami. W dokumencie podkreślono, że standaryzacja formatów danych i wprowadzenie zasad otwartych danych może zapobiec powielaniu informacji, zminimalizować straty czasu i poprawić dokładność prognoz.

W przypadku branży budowlanej, w której tradycyjnie dominuje wysoki stopień fragmentacji i różnorodności formatów, ustrukturyzowany proces unifikacji i ustrukturyzowane otwarte dane odgrywają kluczową rolę w kształtowaniu spójnych i łatwych w zarządzaniu procesów (rys. 4.1-14). Pozwalają one uczestnikom projektu skupić się na poprawie wydajności, a nie na rozwiązywaniu problemów technicznych związanych z niekompatybilnością między zamkniętymi platformami, modelami danych i formatami.



Rys. 3.2-15 Otwarte dane strukturalne zmniejszają zależność od oprogramowania i platform oraz przyspieszają wprowadzanie innowacji.

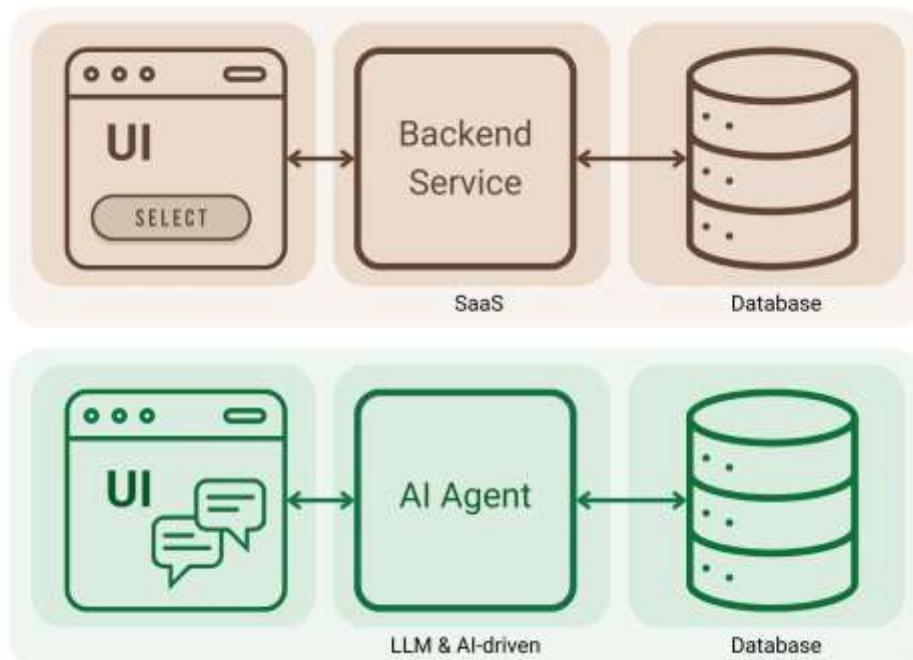
Nowoczesne narzędzia technologiczne, które omówimy szczegółowo w dalszej części książki, pozwalają nie tylko gromadzić informacje, ale także automatycznie je czyścić: eliminować duplikaty, poprawiać błędy i normalizować wartości. Oznacza to, że analitycy i inżynierowie nie pracują z różnymi dokumentami, ale ze zorganizowaną bazą wiedzy odpowiednią do analizy, automatyzacji i podejmowania decyzji.

Zrób to tak prosto, jak to możliwe, ale nie prościej.

- Albert Einstein, fizyk teoretyczny (autorstwo cytatu jest kwestionowane [84])

Obecnie większość interfejsów użytkownika do pracy z danymi może być tworzona automatycznie - bez konieczności ręcznego pisania kodu dla każdego przypadku biznesowego. Wymaga to warstwy infrastruktury, która rozumie strukturę danych, model i logikę bez dodatkowych instrukcji (Rys. 4.1-15). To właśnie ustrukturyzowane dane umożliwiają takie podejście: formularze, tabele, filtry i widoki mogą być generowane automatycznie przy minimalnym wysiłku programistycznym.

Najważniejsze interfejsy krytyczne dla użytkownika mogą nadal wymagać ręcznego dopracowania. Jednak w większości przypadków - a jest to od 50 do 90 procent scenariuszy pracy - automatyczne generowanie aplikacji i obliczeń bez użycia specjalnych aplikacji do tego celu jest wystarczające (rys. 3.2-16), co znacznie obniża koszty rozwoju i utrzymania, zmniejsza liczbę błędów i przyspiesza wdrażanie rozwiązań cyfrowych.



Rys. 3.2-16 Modele architektoniczne do pracy z danymi: tradycyjna architektura aplikacji i model zorientowany na AI z LLM.

Przejście od architektur opartych na pojedynczych aplikacjach do inteligentnie zarządzanych systemów opartych na modelach językowych (LLM) jest kolejnym krokiem w cyfrowej ewolucji. W takiej architekturze ustrukturyzowane dane stają się nie tylko przedmiotem przechowywania, ale także podstawą interakcji z narzędziami sztucznej inteligencji zdolnymi do analizowania, interpretowania i rekomendowania działań w oparciu o kontekst.

W kolejnych rozdziałach przyjrzymy się rzeczywistym przykładom wdrażania architektury opartej na otwartych danych strukturalnych i pokażemy, w jaki sposób modele językowe są stosowane do automatycznej interpretacji, walidacji i przetwarzania danych. Te praktyczne przypadki pomogą Ci lepiej zrozumieć, jak nowa logika cyfrowa działa w akcji - i jakie korzyści przynosi firmom, które są gotowe na transformację.



ROZDZIAŁ 2.3.

LLM I ICH ROLA W PRZETWARZANIU DANYCH I PROCESACH BIZNESOWYCH

LLM czaty: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok do automatyzacji procesów danych

Pojawienie się dużych modeli językowych (LLM) było naturalnym rozszerzeniem ruchu w kierunku ustrukturyzowanych otwartych danych i filozofii Open Source. Kiedy dane stają się zorganizowane, dostępne i czytelne dla maszyn, następnym krokiem jest narzędzie, które może wchodzić w interakcje z tymi informacjami bez konieczności pisania złożonego kodu lub posiadania specjalistycznej wiedzy technicznej.

LLM są bezpośrednim produktem otwartości: dużych otwartych zbiorów danych, publikacji i ruchu Open Source. Bez otwartych artykułów naukowych, publicznie dostępnych danych tekstowych i kultury wspólnego rozwoju nie byłoby ChatGPT ani innych programów LLM. LLM jest w pewnym sensie "destylatem" zgromadzonej przez ludzkość wiedzy cyfrowej, zebranej i wykształconej dzięki zasadom otwartości.

Nowoczesne duże modele językowe (LLM - Large Language Models), takie jak ChatGPT ® (OpenAI), LlaMa™ (Meta AI), Mistral DeepSeek™, Grok™ (xAI), Claude™ (Anthropic), QWEN™ zapewniają użytkownikom możliwość formułowania zapytań do danych w języku naturalnym. Dzięki temu praca z informacjami staje się dostępna nie tylko dla programistów, ale także dla analityków, inżynierów, planistów, menedżerów i innych specjalistów, którzy wcześniej byli dalecy od programowania

LLM (Large Language Model) to sztuczna inteligencja, która jest wyszkolona do rozumienia i generowania tekstu na podstawie ogromnych ilości danych zebranych z całego Internetu. Jest w stanie analizować kontekst, odpowiadać na pytania, angażować się w dialog, pisać teksty i generować kod oprogramowania.

O ile wcześniej wizualizacja, przetwarzanie czy analiza danych wymagały znajomości specjalnego języka programowania: Python, SQL, R czy Scala, a także umiejętności pracy z bibliotekami takimi jak Pandas, Polars czy DuckDB i wieloma innymi, to od 2023 roku sytuacja zmieniła się diametralnie. Teraz użytkownik może po prostu opisać, co chce uzyskać - a model sam wygeneruje kod, wykona go, wyświetli tabelę lub wykres i wyjaśni wynik. Po raz pierwszy od dziesięcioleci rozwój technologii nie podąża ścieżką komplikacji, ale radykalnego uproszczenia i dostępności.

Zasada ta - "przetwarzaj dane za pomocą słów (podpowiedzi)" - wyznaczyła nowy etap w ewolucji pracy z informacją, skutecznie przenosząc tworzenie rozwiązań na jeszcze wyższy poziom abstrakcji. Tak jak kiedyś użytkownicy nie musieli już rozumieć technicznych podstaw Internetu, aby prowadzić sklepy internetowe lub tworzyć strony internetowe za pomocą WordPress, Joomla i innych modułowych systemów open source (a książki pracuje z takimi systemami od 2005 roku, w tym z edukacyjnymi i inżynierskimi platformami internetowymi). - To z kolei doprowadziło do rozkwitu treści cyfrowych i biznesu online - dziś

inżynierowie, analitycy i menedżerowie mogą automatyzować przepływy pracy bez znajomości języków programowania. Jest to ułatwione dzięki potężnym programom LLM - zarówno darmowym, jak i open source, takim jak LLaMA, Mistral, Qwen, DeepSeek i inne - które sprawiają, że zaawansowane technologie są dostępne dla jak najszerzego grona odbiorców.

Duże modele językowe LLM: jak to działa?

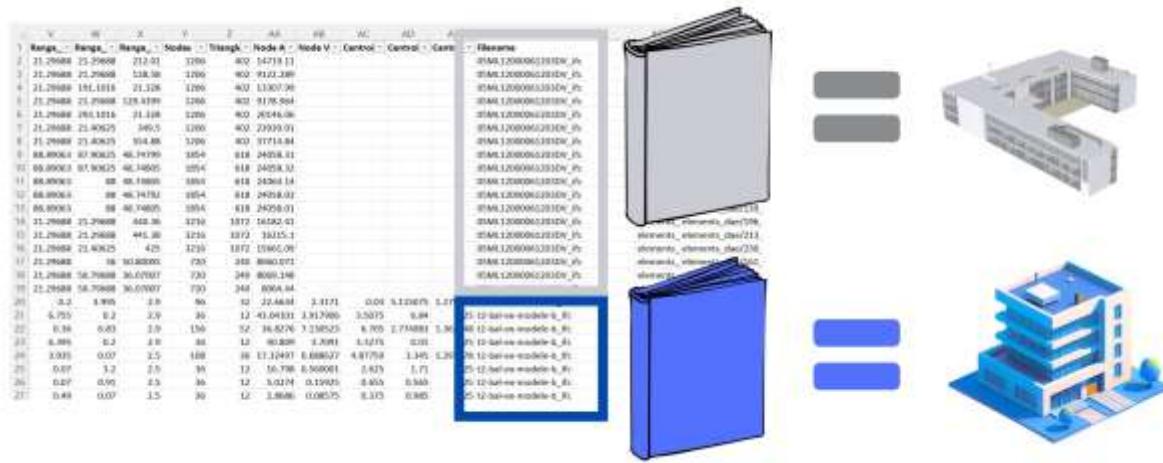
Duże modele językowe (ChatGPT, LLaMA, Mistral, Claude, DeepSeek, QWEN, Grok) to sieci neuronowe trenowane na ogromnych ilościach danych tekstowych z Internetu, książek, artykułów i innych źródeł. Ich głównym zadaniem jest zrozumienie kontekstu ludzkiej mowy i generowanie znaczących odpowiedzi.

Nowoczesny LLM opiera się na architekturze Transformer zaproponowanej przez badaczy Google w 2017 roku [85]. Kluczowym elementem tej architektury jest mechanizm uwagi, który pozwala modelowi uwzględniać relacje między słowami niezależnie od ich pozycji w tekście.

Proces uczenia się LLM jest zdalnie podobny do sposobu, w jaki ludzie uczą się języka, tylko miliony razy większy. Model analizuje miliardy przykładów słów i wyrażeń, identyfikując wzorce w strukturze języka i logice przejść semantycznych. Cały tekst jest dzielony na tokeny - minimalne jednostki semantyczne (słowa lub ich części), które są następnie przekształcane w wektory w wielowymiarowej przestrzeni (rys. 8.2-2). Te reprezentacje wektorowe pozwalają maszynie "zrozumieć" ukryte relacje między pojęciami, zamiast po prostu operować na tekście jako sekwencji symboli.

Modele Big Language to nie tylko narzędzia do generowania tekstu. Są one w stanie rozpoznawać znaczenie, znajdować powiązania między pojęciami i pracować z danymi, nawet jeśli są one prezentowane w różnych formatach. Najważniejsze jest to, że informacje powinny być podzielone na zrozumiałe modele i reprezentowane jako tokeny, z którymi LLM może pracować.

To samo podejście można zastosować do projektów budowlanych. Jeśli pomyślimy o projekcie jako rodzaju tekstu, w którym każdy budynek, element lub konstrukcja jest tokenem, możemy zacząć przetwarzać takie informacje w podobny sposób. Projekty budowlane można porównać do książek, które są zorganizowane w kategorie, rozdziały i grupy akapitów składające się z minimalnych tokenów - elementów projektu budowlanego (Rys. 3.3-1). Tłumacząc modele danych na ustrukturyzowany format, możemy również przetłumaczyć ustrukturyzowane dane na bazy wektorowe (rys. 8.2-2), które są idealnym źródłem dla uczenia maszynowego i technologii takich jak LLM.



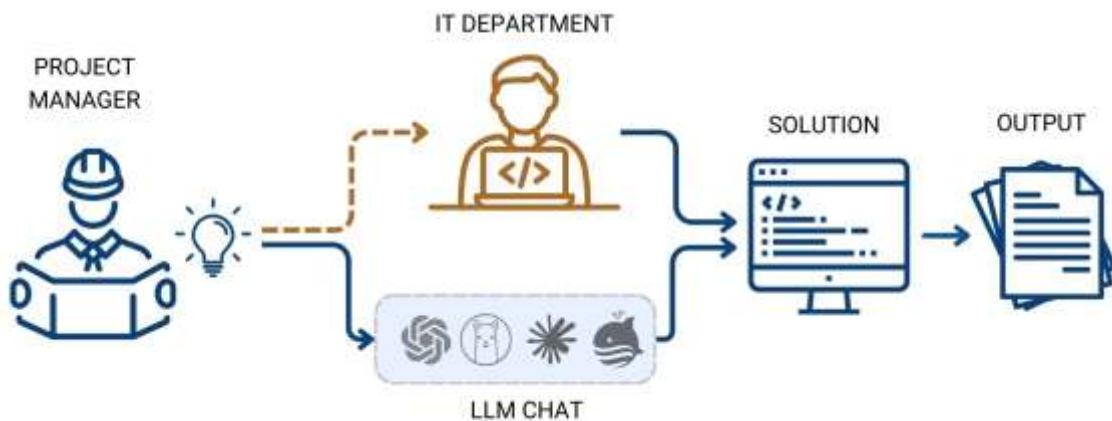
Rys. 3.3-1 Element projektu budowlanego jest jak żeton w tekście: minimalna jednostka, z której tworzone są grupy (akapity) sekcji (kategorii) całego projektu.

Jeśli projekt budowlany jest zdigitalizowany, a jego elementy są reprezentowane jako tokeny lub wektory, możliwy staje się dostęp do nich w języku naturalnym, a nie poprzez sztywne formalne zapytania. W tym miejscu pojawia się jedna z kluczowych zalet LLM - możliwość zrozumienia znaczenia zapytania i powiązania go z odpowiednimi danymi.

Inżynier nie musi już pisać zapytań SQL lub kodu Python, aby uzyskać wymagane dane - może po prostu, rozumiejąc LLM i strukturę danych, sformułować zadanie w zwykły sposób: "Znajdź wszystkie konstrukcje żelbetowe o klasie betonu wyższej niż B30 i oblicz ich całkowitą objętość". Model rozpozna znaczenie zapytania, przekształci je w formę czytelną dla maszyny, znajdzie dane (pogrupuje i przekształci) i zwróci wynik końcowy.

Dokumenty, tabele, modele projektów są konwertowane na reprezentacje wektorowe (osadzanie) i przechowywane w bazie danych. Gdy użytkownik zadaje pytanie, zapytanie jest również konwertowane na wektor, a system znajduje najbardziej odpowiednie dane. Dzięki temu LLM może polegać nie tylko na własnej wyszkolonej wiedzy, ale także na rzeczywistych danych korporacyjnych, nawet jeśli pojawiły się one już po zakończeniu szkolenia modelu.

Jedną z najważniejszych zalet LLM w budownictwie jest możliwość generowania kodu programu. Zamiast przekazywać zadanie techniczne programistie, specjalisci mogą opisać zadanie w języku naturalnym, a model stworzy niezbędny kod, który można wykorzystać (kopując go z czatu) w tworzeniu kodu automatyzacji procesów. LLM -modele pozwalają specjalistom bez głębokiej wiedzy programistycznej przyczynić się do automatyzacji i usprawnienia procesów biznesowych firmy.



Rys. 3.3-2 LLMs zapewniają użytkownikom możliwość pisania kodu i uzyskiwania wyników bez konieczności posiadania umiejętności programowania.

Według badania przeprowadzonego przez Wakefield Research i sponsorowanego przez SAP w 2024 roku [36], w którym przebadano 300 członków kadry kierowniczej wyższego szczebla w firmach o rocznych przychodach w wysokości co najmniej 1 miliarda dolarów w USA: 52% członków kadry kierowniczej wyższego szczebla ufa sztucznej inteligencji w zakresie analizy danych i przedstawiania zaleceń dotyczących podejmowania decyzji. Kolejne 48% używa sztucznej inteligencji do identyfikowania wcześniej nieuwzględnionych zagrożeń, a 47% używa sztucznej inteligencji do sugerowania alternatywnych planów. Ponadto, 40% wykorzystuje sztuczną inteligencję do opracowywania nowych produktów, planowania budżetu i badań rynkowych. Badanie wykazało również pozytywny wpływ sztucznej inteligencji na życie osobiste, przy czym 39% respondentów zgłosiło poprawę równowagi między życiem zawodowym a prywatnym, 38% zgłosiło poprawę zdrowia psychicznego, a 31% zgłosiło niższy poziom stresu.

Jednak mimo całej swojej mocy, LLM pozostają narzędziem, z którego należy korzystać świadomie. Jak każda technologia, mają one swoje ograniczenia. Jednym z najbardziej znanych problemów są tak zwane "halucynacje" - przypadki, w których model z pewnością daje wiarygodną, ale błędą odpowiedź. Dlatego tak ważne jest zrozumienie, jak działa model: jakie dane i modele danych może bezbłędnie interpretować, jak interpretuje zapytania i skąd czerpie informacje. Warto również pamiętać, że wiedza LLM jest ograniczona do daty jego szkolenia, a bez połączenia z danymi zewnętrznymi model może nie uwzględniać aktualnych norm, standardów, cen lub technologii.

Rozwiązaniem tych problemów jest regularne aktualizowanie wektorowych baz danych, łączenie się z odpowiednimi źródłami i rozwijanie autonomicznej sztucznej inteligencji - agentów, którzy nie tylko odpowiadają na pytania, ale proaktywnie wykorzystują dane do szkolenia, zarządzania zadaniami, identyfikowania ryzyka, sugerowania opcji optymalizacji i monitorowania wydajności projektu.

Przejście na LLM -interfejsy w budownictwie to nie tylko nowość technologiczna. To zmiana paradygmatu, usuwająca bariery między ludźmi a danymi. To możliwość pracy z informacjami tak łatwo, jak rozmawiamy ze sobą - i nadal uzyskiwanie dokładnych, zweryfikowanych i praktycznych wyników.

Firmy, które zaczną korzystać z takich narzędzi wcześniej niż inne, zyskają znaczącą przewagę

konkurencyjną. Obejmuje to przyspieszenie pracy, obniżenie kosztów i poprawę jakości rozwiązań projektowych dzięki szybkiemu dostępowi do analizy danych i możliwości szybkiego znajdowania odpowiedzi na złożone pytania. Należy jednak wziąć pod uwagę również kwestie bezpieczeństwa. Korzystanie z opartych na chmurze usług LLM może wiązać się z ryzykiem wycieku danych. Dlatego organizacje coraz częściej poszukują alternatywnych rozwiązań, które pozwolą im wdrożyć narzędzia LLM we własnej infrastrukturze - lokalnie, z pełną ochroną i kontrolą nad informacjami.

Korzystanie z lokalnych usług LLM dla wrażliwych danych firmowych

Pojawienie się pierwszych chat-LLMs w 2022 r. wyznaczyło nowy etap w rozwoju sztucznej inteligencji. Jednak natychmiast po powszechnym przyjęciu tych modeli pojawiło się uzasadnione pytanie: jak bezpieczne jest przesyłanie danych i zapytań związanych z firmą do chmury? Większość modeli językowych opartych na chmurze przechowywała historię komunikacji i przesłane dokumenty na swoich serwerach, a dla firm zajmujących się wrażliwymi informacjami była to poważna bariera dla przyjęcia sztucznej inteligencji.

Jednym z najbardziej zrównoważonych i logicznych rozwiązań tego problemu było wdrożenie Open Source LLM lokalnie, w ramach korporacyjnej infrastruktury IT. W przeciwieństwie do usług w chmurze, modele lokalne działają bez połączenia z Internetem, nie przesyłają danych na zewnętrzne serwery i dają firmom pełną kontrolę nad informacjami

Najlepszy otwarty model [Open Source LLM] jest obecnie porównywalny pod względem wydajności do modeli zamkniętych [takich jak ChatGPT, Claude], ale z około rocznym opóźnieniem [77].

- Ben Cottier, główny badacz w Epoch AI, organizacji badawczej non-profit, 2024 r.

Duże firmy technologiczne zaczęły udostępniać swoje rozwiązania LLM do użytku lokalnego. Przykładami przejścia na otwartą architekturę były seria LLaMA firmy Meta oraz szybko rozwijający się chiński projekt DeepSeek. Oprócz nich, Mistral i Falcon również wydały potężne modele wolne od ograniczeń zastrzeżonych platform. Inicjatywy te nie tylko przyspieszyły rozwój globalnej sztucznej inteligencji, ale także dały firmom dbającym o prywatność realne alternatywy w zakresie niezależności, elastyczności i zgodności z wymogami bezpieczeństwa.

W środowisku korporacyjnym, zwłaszcza w branży budowlanej, ochrona danych nie jest tylko kwestią wygody, ale zgodności z przepisami. Praca z dokumentami przetargowymi, kosztorysami, rysunkami i poufną korespondencją wymaga ścisłej kontroli. I właśnie w tym obszarze lokalna firma LLM zapewnia niezbędną pewność, że dane pozostają w granicach firmy.

	Cloud LLMs (OpenAI, Claude)	Local LLMs (DeepSeek, LLaMA)
Data Control	Data is transmitted to third parties	Data remains within the company's network
License	Proprietary, paid	Open-source (Apache 2.0, MIT)
Infrastructure	Requires internet	Operates in an isolated environment
Customization	Limited	Full adaptation to company needs
Cost	Pay-per-token/request	One-time hardware investment + maintenance costs
Scalability	Easily scalable with cloud resources	Scaling requires additional local hardware
Security & Compliance	Risk of data leaks, may not meet strict regulations (GDPR, HIPAA)	Full compliance with internal security policies
Performance & Latency	Faster inference due to cloud infrastructure	Dependent on local hardware, may have higher latency
Integration	API-based integration, requires internet access	Can be tightly integrated with on-premise systems
Updates & Maintenance	Automatically updated by provider	Requires manual updates and model retraining
Energy Consumption	Energy cost is covered by provider	High power consumption for inference and training
Offline Availability	Not available without an internet connection	Works completely offline
Inference Cost	Pay-per-use model (cost scales with usage)	Fixed cost after initial investment

Rys. 3.3-3 Modele lokalne zapewniają pełną kontrolę i bezpieczeństwo, podczas gdy rozwiązania oparte na chmurze oferują łatwą integrację i automatyczne aktualizacje.

Kluczowe korzyści z lokalnego Open Source LLM:

- Pełna kontrola nad danymi. Wszystkie informacje pozostają w firmie, eliminując nieautoryzowany dostęp i wyciek danych.
- Praca w trybie offline. Brak zależności od połączenia internetowego, co jest szczególnie ważne w przypadku odizolowanych infrastruktur IT. Zapewnia to również nieprzerwane działanie w obliczu sankcji lub zablokowanych usług w chmurze.
- Elastyczność zastosowań. Model może być wykorzystywany do generowania tekstu, analizy danych, pisania kodu programu, wspomagania projektowania i zarządzania procesami biznesowymi.
- Dostosowanie do celów korporacyjnych. LLM może być szkolony na dokumentach wewnętrznych,

co pozwala na uwzględnienie specyfiki pracy firmy i jej cech branżowych. Lokalny LLM można połączyć z CRM, ERP lub platformami BI, co pozwala zautomatyzować analizę zapytań klientów, tworzenie raportów, a nawet prognozowanie trendów.

Wdrożenie darmowego i otwartego modelu DeepSeek -R1-7B na serwerze, w celu uzyskania dostępu przez cały zespół użytkowników, w cenie 1000 USD miesięcznie może potencjalnie kosztować mniej niż roczne opłaty za interfejsy API w chmurze, takie jak ChatGPT lub Claude i pozwala firmom przejąć pełną kontrolę nad swoimi danymi, eliminując ich przesyłanie do Internetu i pomaga zachować zgodność z wymogami regulacyjnymi, takimi jak RODO

W innych branżach lokalne LLM już zmieniają swoje podejście do automatyzacji. W usługach wsparcia odpowiadają na częste zapytania klientów, zmniejszając obciążenie operatorów. W działach HR analizują CV i wybierają odpowiednich kandydatów. W handlu elektronicznym generują spersonalizowane oferty bez ujawniania danych użytkowników.

Podobny efekt spodziewany jest w sektorze budowlanym. Dzięki integracji LLM z danymi projektowymi i standardami, możliwe jest przyspieszenie przygotowania dokumentacji, automatyzacja przygotowania szacunków i predykcyjnych analiz kosztów. Wykorzystanie LLM w połączeniu z ustrukturyzowanymi tabelami i ramami danych staje się szczególnie obiecującym obszarem.

Pełna kontrola nad AI w firmie i jak wdrożyć własne LLM

Nowoczesne narzędzia pozwalają firmom wdrożyć duży model językowy (LLM) lokalnie w ciągu zaledwie kilku godzin. Daje to pełną kontrolę nad danymi i infrastrukturą, eliminując zależność od zewnętrznych usług w chmurze i minimalizując ryzyko wycieku informacji. Rozwiązywanie to jest szczególnie istotne dla organizacji pracujących z wrażliwą dokumentacją projektową lub poufnymi danymi handlowymi.

W zależności od zadań i zasobów dostępne są różne scenariusze wdrażania, od gotowych rozwiązań po bardziej elastyczne i skalowalne architektury. Jednym z najprostszych narzędzi jest Ollama, która umożliwia uruchamianie modeli językowych dosłownie jednym kliknięciem, bez konieczności posiadania głębokiej wiedzy technicznej. Szybki start z Ollamą:

1. Pobierz dystrybucję dla swojego systemu operacyjnego (Windows / Linux / macOS) z oficjalnej strony internetowej: ollama.com
2. Zainstaluj model za pomocą wiersza poleceń. Na przykład dla modelu *Mistral*:

```
ollama run mistral
```

3. Po uruchomieniu modelu jest on gotowy do pracy - można wysyłać zapytania tekstowe za pośrednictwem terminala lub zintegrować go z innymi narzędziami. Uruchom model i wykonaj zapytanie:

ollama run mistral "Jak stworzyć kalkulację ze wszystkimi zasobami do pracy, aby zainstalować ściankę działową z płyt gipsowo-kartonowych o szerokości 100 mm?".

Dla tych, którzy wolą pracować w znajomym środowisku wizualnym, dostępna jest darmowa aplikacja LM Studio z interfejsem przypominającym ChatGPT

- Zainstaluj LM Studio, pobierając zestaw dystrybucyjny z oficjalnej strony internetowej - lmstudio.ai
- Za pomocą wbudowanego katalogu wybierz model (np. Falcon lub GPT-Neo-X) i pobierz go.
- Pracuj z modelem za pomocą intuicyjnego interfejsu przypominającego ChatGPT, ale całkowicie zlokalizowanego.

	Developer	Parameters	GPU Requirements (GB)	Features	Best For
Mistral 7B	Mistral AI	7	8 (FP16)	Fast, supports multimodal tasks (text + images), fully open-source code	Lightweight tasks, mobile devices, laptops
LLaMA 2	Meta	7–70	16–48 (FP16)	High text generation accuracy, adaptable for technical tasks, CC-BY-SA license	Complex analytical and technical tasks
Baichuan 7B/13B	Baichuan Intelligence	7–13	8–16 (FP16)	Fast and efficient, great for large data processing, fully open-source code	Data processing, automating routine tasks
Falcon 7B/40B	Technology Innovation Institute (TII)	7–40	8–32 (FP16)	Open-source, high performance, optimized for fast work	Workloads with limited computational resources
DeepSeek-V3	DeepSeek	671	1543 (FP16) / 386 (4-bit)	Multilingual, 128K token context window, balanced speed and accuracy	Large enterprises, SaaS platforms, multitasking scenarios
DeepSeek-R1-7B	DeepSeek	7	18 (FP16) / 4.5 (4-bit)	Retains 92% of R1 capabilities in MATH-500, local deployment support	Budget solutions, IoT devices, edge computing

Rys. 3.3-4 Porównanie popularnych lokalnych modeli open source LLM.

Wybór modelu zależy od wymagań dotyczących szybkości, dokładności i dostępnych możliwości sprzętowych (Rys. 3.3-4). Małe modele, takie jak Mistral 7B i Baichuan 7B, są odpowiednie do lekkich zadań i urządzeń mobilnych, podczas gdy potężne modele, takie jak DeepSeek -V3, wymagają znacznych zasobów obliczeniowych, ale oferują wysoką wydajność i obsługę wielu języków. W nadchodzących latach rynek LLM będzie się szybko rozwijał - będziemy widzieć coraz więcej lekkich i wyspecjalizowanych modeli. Zamiast modeli LLM ogólnego przeznaczenia obejmujących wszystkie ludzkie treści, pojawią się modele

przeszkolone w zakresie wąskiej wiedzy specjalistycznej. Na przykład, możemy spodziewać się pojawienia się modeli zaprojektowanych wyłącznie do obsługi obliczeń inżynierijnych, szacunków budowlanych lub danych w formatach CAD. Takie wyspecjalizowane modele będą szybsze, dokładniejsze i bezpieczniejsze w użyciu - zwłaszcza w środowiskach zawodowych, w których ważna jest wysoka niezawodność i głęboka znajomość tematu.

Po uruchomieniu lokalnego modelu LLM można go dostosować do konkretnych zadań firmy. W tym celu stosuje się technikę dostrajania, w ramach której model jest dalej szkolony w zakresie dokumentów wewnętrznych, instrukcji technicznych, szablonów umów lub dokumentacji projektowej.

RAG: Inteligentny LLM - asystenci z dostępem do danych korporacyjnych

Kolejnym etapem ewolucji zastosowań LLM w biznesie jest integracja modeli z rzeczywistymi danymi korporacyjnymi w czasie rzeczywistym. Podejście to nosi nazwę RAG (Retrieval-Augmented Generation) - pobieranie-rozszerzone generowanie. W tej architekturze model językowy staje się nie tylko interfejsem dialogowym, ale pełnoprawnym inteligentnym asystentem zdolnym do nawigowania po dokumentach, rysunkach, bazach danych i dostarczania dokładnych, kontekstowych odpowiedzi.

Główną zaletą RAG jest możliwość wykorzystania wewnętrznych danych firmy bez konieczności wstępnego trenowania modelu, przy jednoczesnym zachowaniu wysokiej dokładności i elastyczności w przetwarzaniu informacji.

Technologia RAG łączy w sobie dwa główne elementy:

- **Pobieranie:** model łączy się z magazynami danych - dokumentami, tabelami, plikami PDF, rysunkami - i pobiera odpowiednie informacje zgodnie z żądaniem użytkownika.
- **Rozszerzone generowanie:** na podstawie wyodrębnionych danych model generuje dokładną, świadomą odpowiedź, biorąc pod uwagę kontekst i specyfikę zapytania.

Aby uruchomić LLM z obsługą RAG, należy wykonać kilka kroków:

- **Przygotowanie danych:** zebranie niezbędnych dokumentów, rysunków, specyfikacji, tabel. Mogą one być w różnych formatach i strukturach, od PDF do Excel.
- **Indeksowanie i wektoryzacja:** za pomocą narzędzi takich jak LlamaIndex lub LangChain, dane są konwertowane na reprezentacje wektorowe, które pozwalają znaleźć powiązania semantyczne między fragmentami tekstu (więcej o wektorowych bazach danych i tłumaczeniu dużych tablic na reprezentację wektorową, w tym projektów CAD, w części 8).
- **Zadawanie pytań asystentowi:** po przesłaniu danych możesz zadawać modelowi pytania, a on będzie szukać odpowiedzi w ramach korporacyjnych, a nie w ogólnej wiedzy zebranej z Internetu.

Założymy, że firma ma folder constructionsite_docs, w którym przechowywane są umowy, instrukcje, szacunki i tabele. Używając skryptu Pythona (rys. 3.3-5), możemy przeskanować ten folder i zbudować indeksowanie wektorowe: każdy dokument zostanie przekształcony w zestaw wektorów odzwierciedlających semantyczną zawartość tekstu. W ten sposób dokumenty stają się swego rodzaju "mapą

znaczeń", po której model może sprawnie nawigować i znajdować połączenia między terminami i frazami.

Na przykład, model "pamięta", że słowa "zwrot" i "reklamacja" często znajdują się w części umowy dotyczącej wysyłki materiałów na plac budowy. Następnie, jeśli zostanie zadane pytanie - na przykład "Jaki jest nasz okres zwrotu?" (Rys. 3.3-5 - linia 11 kodu) - LLM przeanalizuje wewnętrzne dokumenty i znajdzie dokładne informacje, działając jak inteligentny asystent zdolny do czytania i rozumienia zawartości wszystkich plików korporacyjnych.



```

1  from llama_index import SimpleDirectoryReader, VectorStoreIndex
2
3  # Load documents from the folder
4  documents = SimpleDirectoryReader("constructionsite_docs").load_data()
5
6  # Creating a vector index for semantic search
7  index = VectorStoreIndex.from_documents(documents)
8
9  # Integration with LLM (e.g. Llama 3)
10 query_engine = index.as_query_engine()
11 response = query_engine.query("What are the return terms in the contracts?")
12 print(response)

```

Rys. 3.3-5 LM odczytuje folder z plikami - podobnie jak człowiek otwiera go i wyszukuje żądzany dokument

Kod można uruchomić na dowolnym komputerze z zainstalowanym Pythonem. Więcej o używaniu Pythona i IDE do uruchamiania kodu powiemy w następnym rozdziale.

Lokalne wdrażanie LLM to nie tylko trend, ale strategiczne rozwiązanie dla firm, które cenią sobie bezpieczeństwo i elastyczność. Jednak wdrożenie LLM, czy to na lokalnych komputerach firmowych, czy przy użyciu rozwiązań online, to dopiero pierwszy krok. Aby zastosować możliwości LLM do rzeczywistych zadań, firmy muszą korzystać z narzędzi, które pozwalają im nie tylko odbierać odpowiedzi na czacie, ale także przechowywać utworzoną logikę w postaci kodu, który można uruchomić poza kontekstem korzystania z LLM. Jest to ważne dla skalowania rozwiązań - odpowiednio zorganizowane procesy pozwalają na zastosowanie rozwoju AI w kilku projektach jednocześnie lub nawet w całej firmie.

W tym kontekście ważną rolę odgrywa wybór odpowiedniego środowiska programistycznego (IDE). Nowoczesne narzędzia programistyczne pozwalają nie tylko tworzyć rozwiązania oparte na LLM, ale także integrować je z istniejącymi procesami biznesowymi, przekształcając je w zautomatyzowane ETL - Pipeline



ROZDZIAŁ 2.4.

IDE Z OBSŁUGĄ LLM I PRZYSZŁYMI ZMIANAMI W PROGRAMOWANIU

Wybór IDE: od eksperymentów LLM do rozwiązań biznesowych

Zanurzając się w świat automatyzacji, analizy danych i sztucznej inteligencji - zwłaszcza podczas pracy z dużymi modelami językowymi (LLM) - kluczowe znaczenie ma wybór odpowiedniego zintegrowanego środowiska programistycznego (IDE). To IDE będzie głównym narzędziem pracy: miejscem, w którym kod generowany przez LLM będzie uruchamiany, zarówno na komputerze lokalnym, jak i w sieci korporacyjnej. Wybór IDE decyduje nie tylko o wygodzie pracy, ale także o tym, jak szybko będziesz w stanie przejść od eksperymentalnych zapytań LLM do pełnoprawnych rozwiązań osadzonych w rzeczywistych procesach biznesowych.

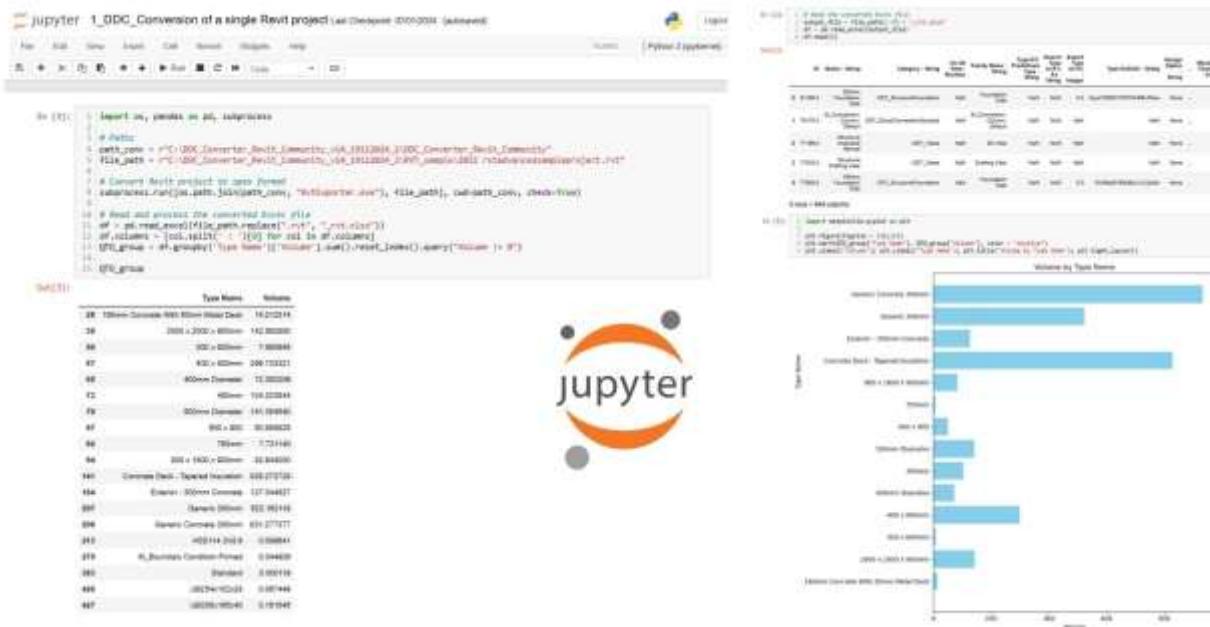
IDE (Integrated Development Environment) to wszechstronne narzędzie do automatyzacji procesów i przetwarzania danych. Zamiast trzymać osobno piłę, młotek, wiertarkę i inne narzędzia, masz jedno urządzenie, które potrafi wszystko - ciąć, mocować, wiercić, a nawet sprawdzać jakość materiałów. IDE dla programistów to jedna przestrzeń, w której można pisać kod (analogicznie do budowy - tworzyć rysunki), testować jego działanie (montaż modelu budynku), znajdować błędy (jak sprawdzanie wytrzymałości konstrukcji w budownictwie) i uruchamiać gotowy projekt (oddanie domu do użytku).

Przegląd popularnych IDE:

- **PyCharm®** (JetBrains) to potężne, profesjonalne IDE dla Pythona. Ze względu na dużą liczbę wbudowanych funkcji doskonale nadaje się do poważnych projektów. Jednak podstawowa obsługa interaktywnych plików Jupyter (IPYNB) jest dostępna tylko w wersji płatnej, a początkujący mogą uznać interfejs za przytłaczający.

Plik z rozszerzeniem IPYNB (Interactive Python Notebook) to format interaktywnych notatników Jupyter® (Rys. 3.4-1), w których kod, wizualizacje i objaśnienia są połączone w jednym dokumencie. Format ten jest idealny do tworzenia raportów, analiz i scenariuszy szkoleniowych.

- **VS Code®** (Microsoft) to szybkie, elastyczne i konfigurowalne narzędzie z bezpłatną obsługą IPYNB i wieloma wtyczkami. Odpowiednio zarówno dla początkujących, jak i profesjonalistów. Umożliwia integrację wtyczek GitHub Copilot i modelu językowego, dzięki czemu jest doskonałym wyborem dla projektów AI i data science.
- **Jupyter Notebook** - klasyczny i popularny wybór do eksperymentowania i nauki. Umożliwia pisanie kodu, dodawanie wyjaśnień i wizualizację wyników w jednym interfejsie (rys. 3.4-1). Idealny do szybkiego testowania hipotez, pracy z LLM i tworzenia powtarzalnych dzikich etapów analizy danych. Do zarządzania zależnościami i bibliotekami zalecamy korzystanie z Anaconda Navigator, wizualnego interfejsu do zarządzania środowiskiem Python.



Rys. 3.4-1 Jupyter Notebook jedno z najwygodniejszych i najpopularniejszych narzędzi do tworzenia procesów Pipeline.

■ **Google Collab™** (i platforma Kaggle (Rys. 9.2-5)) to oparta na chmurze alternatywa dla Jupytera, która zapewnia bezpłatny dostęp do GPU/TPU. Jest to świetne rozwiązanie na początek - brak konieczności instalacji lokalnego oprogramowania i możliwość pracy bezpośrednio z przeglądarki. Obsługuje integrację z Dyskiem Google, a ostatnio także z Gemini (LLM Google).

	PyCharm	VS Code	Jupyter Notebook	Google Colab
Complexity	High	Medium	Low	Low
.ipynb support	Paid	Free	Built-in	Built-in
Copilots	Yes	Yes	Yes	Yes
Computing resources	Local	Local	Local	Cloud
For whom	Professionals	Universal	Beginners	Experimenters

Rys. 3.4-2 Porównanie IDE: Jupyter Notebook jedno z najwygodniejszych i najłatwiejszych narzędzi do tworzenia procesów Pipeline.

Wybór IDE zależy od wykonywanych zadań. Jeśli chcesz szybko rozpoczęć pracę z AI, wypróbuj Jupyter Notebook lub Google Collab. W przypadku poważnych projektów lepiej jest użyć PyCharm lub VS Code. Najważniejsze to zacząć. Nowoczesne narzędzia pozwalają szybko przekształcić eksperymenty w działające rozwiązania.

Wszystkie opisane powyżej IDE pozwalały na tworzenie potoków przetwarzania danych - czyli łańcuchów bloków kodu (które mogą być generowane przez LLM), z których każdy odpowiada np. za inny etap:

- scenariusze analityczne,
- łańcuchy ekstrakcji informacji z dokumentów,
- automatyczne reakcje na podstawie RAG,
- generowanie raportów i wizualizacji.

Dzięki modułowej strukturze, każdy krok może być reprezentowany jako oddzielny blok: ładowanie danych → filtrowanie → analiza → wizualizacja → eksportowanie wyników. Bloki te mogą być ponownie wykorzystane, - dostosowane i złożone w nowe łańcuchy, jak konstruktor, tylko dla danych.

Dla inżynierów, menedżerów i analityków otwiera to możliwość dokumentowania logiki decyzyjnej w postaci kodu, który można wygenerować za pomocą LLM. Takie podejście pomaga przyspieszyć rutynowe zadania, zautomatyzować typowe operacje i stworzyć powtarzalne procesy, w których każdy krok jest jasno udokumentowany i przejrzysty dla wszystkich członków zespołu.

Zautomatyzowane narzędzia ETL Pipelines (rys. 7.2-3), Apache Airflow (rys. 7.4-4), Apache NiFi (rys. 7.4-5) i n8n (rys. 7.4-6) do budowania bloków logiki do automatyzacji procesów zostaną omówione bardziej szczegółowo w części 7 i części 8 książki.

IDE z obsługą LLM i przyszłe zmiany w programowaniu

Integracja sztucznej inteligencji z procesami programistycznymi zmienia krajobraz programowania. Nowoczesne środowiska nie są już tylko edytormi tekstu z podświetlaniem składni - zmieniają się w intelligentnych asystentów zdolnych do zrozumienia logiki projektu, uzupełnienia kodu, a nawet wyjaśnienia, jak działa dany fragment kodu. Na rynku pojawiają się produkty, które wykorzystują sztuczną inteligencję do przesuwania granic konwencjonalnego rozwoju:

- **GitHub Copilot** (integruje się z VS Code, PyCharm): asystent AI, który generuje kod na podstawie komentarzy lub częściowych opisów, zamieniając tekstowe wskazówki w gotowe rozwiązania.
- **Cursor** (fork VS Code z AI -kernel): pozwala nie tylko dokończyć pisanie kodu, ale także zadawać pytania do projektu, szukać zależności i uczyć się z bazy kodu.
- **JetBrains AI Assistant**: wtyczka do IDE JetBrains (w tym PyCharm) z funkcją wyjaśniania złożonego kodu, optymalizacji i tworzenia testów.
- **Amazon CodeWhisperer**: odpowiednik Copilot z naciskiem na bezpieczeństwo i wsparcie dla usług AWS firmy Amazon.

W nadchodzących latach programowanie przejdzie dramatyczną zmianę. Główny nacisk zostanie przeniesiony z rutynowego pisania kodu na projektowanie modeli i architektury danych - programiści będą bardziej zaangażowani w projektowanie systemu, podczas gdy sztuczna inteligencja przejmie zadania szablonowe: generowanie kodu, testy, dokumentację i podstawowe funkcje. Przyszłość programowania to współpraca między ludźmi a sztuczną inteligencją, w której maszyny przejmują rutynę techniczną, a ludzie koncentrują się na kreatywności.

Programowanie w języku naturalnym stanie się codziennością. Personalizacja IDE osiągnie nowy poziom - środowiska programistyczne nauczą się dostosowywać do stylu pracy użytkownika i firm, przewidując

wzorce, oferując rozwiązania kontekstowe i ucząc się na podstawie poprzednich projektów

Nie zlikwiduje to roli programisty, ale radykalnie ją przekształci: od pisania kodu do zarządzania wiedzą, jakością i procesami. Ewolucja ta wpłynie również na analitykę biznesową, w której tworzenie raportów, wizualizacji i aplikacji wspomagających podejmowanie decyzji będzie w coraz większym stopniu odbywać się poprzez generowanie kodu i logiki za pomocą sztucznej inteligencji i LLM, czatu i interfejsów agentów.

Po skonfigurowaniu czatów LLM i wybraniu odpowiedniego środowiska programistycznego, kolejnym ważnym krokiem jest uporządkowanie danych. Proces ten obejmuje wyodrębnianie informacji z różnych źródeł, czyszczenie ich, przekształcanie w ustrukturyzowaną formę i integrowanie z systemami korporacyjnymi.

W nowoczesnym podejściu Data-Centric do zarządzania danymi, kluczowym celem jest doprowadzenie danych do jednej uniwersalnej formy, która jest kompatybilna z dużą liczbą narzędzi i aplikacji. Do obsługi procesów strukturyzacji i ustrukturyzowanych danych potrzebne są wyspecjalizowane biblioteki. Jedną z najbardziej wydajnych, elastycznych i popularnych jest biblioteka Pandas dla Pythona. Umożliwia ona wygodne przetwarzanie danych tabelarycznych: filtrowanie, grupowanie, czyszczenie, dołączanie, agregowanie i raportowanie.

Python Pandas: niezastąpione narzędzie do pracy z danymi

Pandas zajmuje szczególne miejsce w świecie analizy danych i automatyzacji. Jest to jedna z najpopularniejszych i najczęściej używanych bibliotek języka programowania Python [86], zaprojektowana do pracy z danymi strukturalnymi.

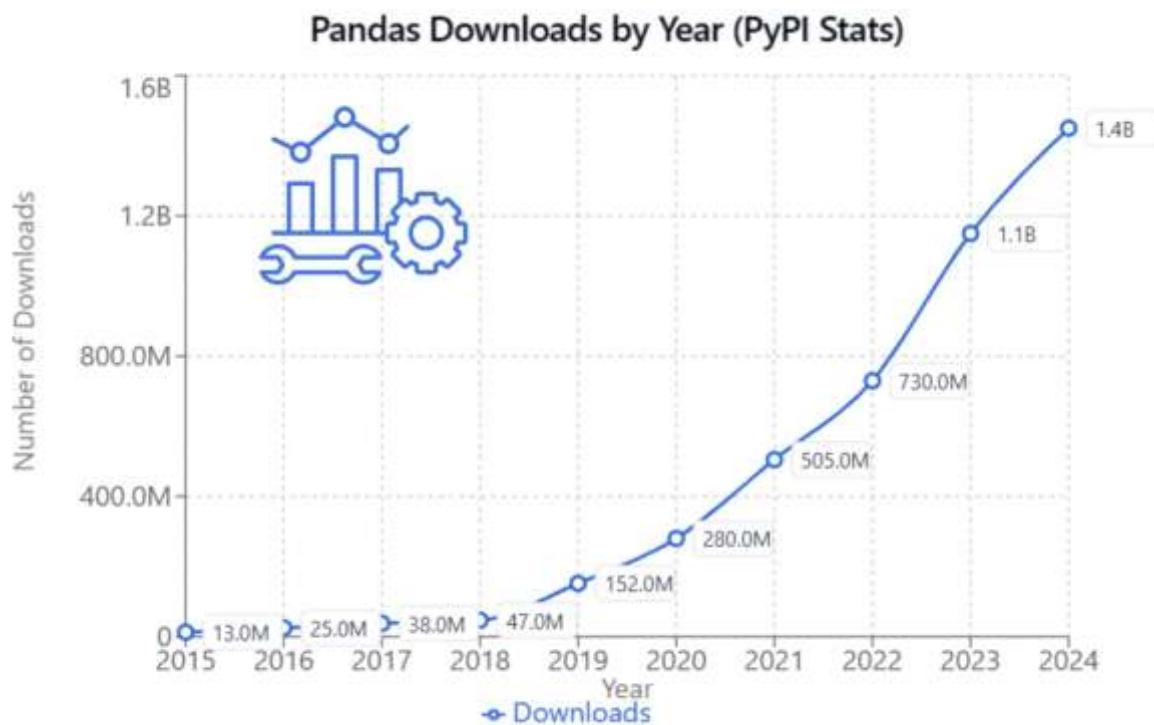
Biblioteka jest jak zestaw gotowych narzędzi: funkcji, modułów, klas. Tak jak na budowie nie trzeba za każdym razem wymyślać młotka czy poziomicy, tak w programowaniu biblioteki pozwalają na szybkie rozwiązywanie problemów bez wymyślania na nowo podstawowych funkcji i rozwiązań.

Pandas to biblioteka Pythona o otwartym kodzie źródłowym, zapewniająca wysoką wydajność i intuicyjne struktury danych, w szczególności DataFrame, uniwersalny format do pracy z tabelami. Pandas to szwajcarski scyzoryk dla analityków, inżynierów i programistów pracujących z danymi.

Python to język programowania wysokiego poziomu o prostej składni, który jest aktywnie wykorzystywany w analityce, automatyzacji, uczeniu maszynowym i tworzeniu stron internetowych. Jego popularność wynika z czytelności kodu, wieloplatformowości i bogatego ekosystemu bibliotek. Do tej pory stworzono ponad 137 000 pakietów open source dla Pythona [87], a liczba ta rośnie niemal codziennie. Każda taka biblioteka jest swego rodzaju repozytorium gotowych funkcji: od prostych operacji matematycznych po złożone narzędzia do przetwarzania obrazu, analizy dużych zbiorów danych, sieci neuronowych i integracji z zewnętrznymi usługami.

Innymi słowy, wyobraź sobie, że masz bezpłatny i otwarty dostęp do setek tysięcy gotowych rozwiązań programistycznych - bibliotek i narzędzi, które możesz bezpośrednio osadzić w swoich procesach biznesowych. To jak ogromny katalog aplikacji do automatyzacji, analizy, wizualizacji, integracji i nie tylko - a wszystko to dostępne natychmiast po zainstalowaniu Pythona.

Pandas jest jednym z najpopularniejszych pakietów w ekosystemie Pythona. W 2022 r. średnia liczba pobrań biblioteki Pandas osiągnęła 4 miliony dziennie (Rys. 3.4-3), podczas gdy na początku 2025 r. liczba ta wzrosła do 12 milionów pobrań dziennie, odzwierciedlając jej rosnącą popularność i szerokie zastosowanie w analityce danych i czacie LLM [86]



Rys. 3.4-3 Pandas jest jedną z najczęściej pobieranych bibliotek. W 2024 r. roczna liczba pobrań przekroczyła 1,4 miliarda.

Język zapytań w bibliotece Pandas jest podobny pod względem funkcjonalności do języka zapytań SQL, który omówiliśmy w rozdziale "Relacyjne bazy danych i język zapytań SQL".

W świecie analityki i zarządzania danymi strukturalnymi, Pandas wyróżnia się prostotą, szybkością i mocą, zapewniając użytkownikom szeroką gamę narzędzi do efektywnej analizy i przetwarzania informacji.

Oba narzędzia - SQL i Pandas - zapewniają potężne możliwości manipulacji danymi, zwłaszcza w

porównaniu z tradycyjnym Exceliem. Obsługują operacje takie jak selekcja, filtrowanie (Rys. 3.4-4), z tą różnicą, że SQL jest zoptymalizowany do pracy z relacyjnymi bazami danych, podczas gdy Pandas przetwarza dane w pamięci RAM, co pozwala mu działać na dowolnym komputerze, bez konieczności tworzenia baz danych i wdrażania oddzielnej infrastruktury.

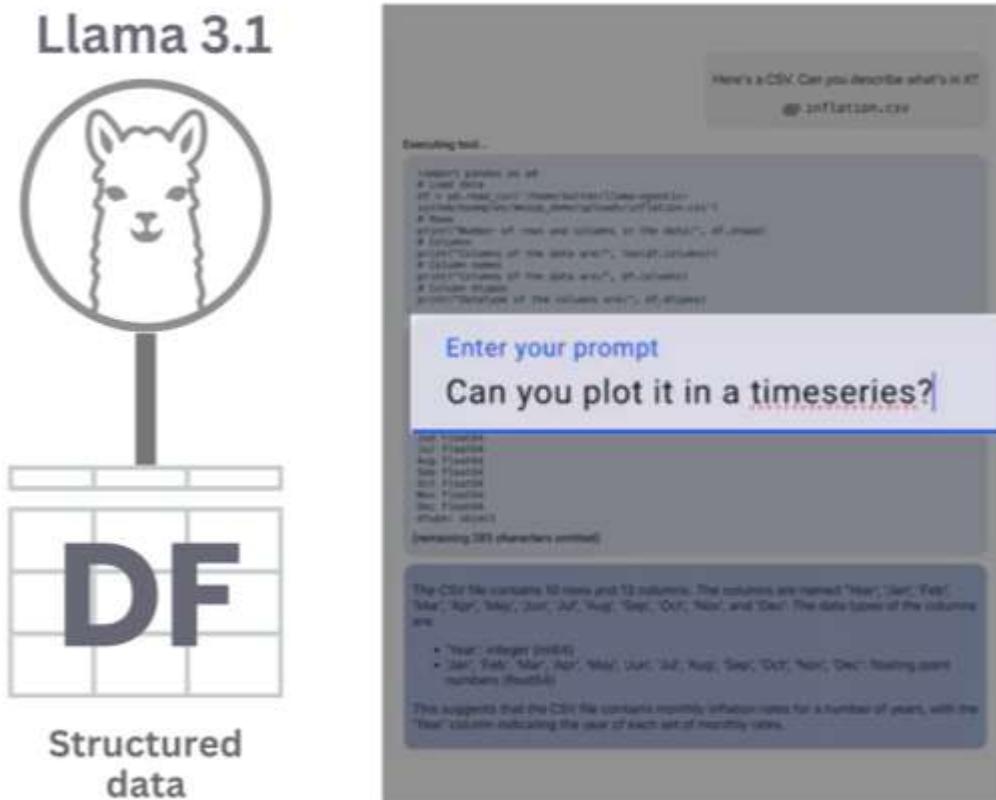


Rys. 3.4-4 Pandas, w przeciwieństwie do SQL, ma elastyczność do pracy z różnymi formatami danych, nie ograniczając się do baz danych.

Pandas jest często preferowany do badań naukowych, automatyzacji procesów, tworzenia potoków (w tym ETL) i manipulacji danymi w Pythonie, podczas gdy SQL jest standardem zarządzania bazami danych i jest często używany w środowiskach korporacyjnych do obsługi dużych ilości danych.

Biblioteka Pandas języka programowania Python umożliwia wykonywanie nie tylko podstawowych operacji, takich jak odczytywanie i zapisywanie tabel, ale także bardziej złożonych zadań, w tym scalanie danych, grupowanie danych i wykonywanie złożonych obliczeń analitycznych.

Obecnie biblioteka Pandas jest wykorzystywana nie tylko w badaniach akademickich i analityce biznesowej, ale także w połączeniu z modelami LLM. Na przykład dział Meta® (Facebook™), publikując nowy model open source LLaMA 3.1 w 2024 r., zwrócił szczególną uwagę na pracę z danymi strukturalnymi, czyniąc jednym z kluczowych i pierwszych przypadków w jego wydaniu właśnie przetwarzanie ustrukturyzowanych ramek danych (rys. 3.4-5) w formacie CSV i integrację z biblioteką Pandas bezpośrednio na czacie.



Rys. 3.4-5 Jednym z pierwszych i głównych przypadków zespołu Meta przedstawionych w LlaMa 3.1 w 2024 r. było tworzenie aplikacji przy użyciu Pandas.

Pandas jest niezbędnym narzędziem dla milionów naukowców przetwarzających i przygotowujących dane do generatywnej sztucznej inteligencji. Przyspieszenie Pandas przy zerowych zmianach kodu będzie ogromnym krokiem naprzód. Naukowcy zajmujący się danymi będą mogli przetwarzać dane w ciągu kilku minut zamiast godzin i uzyskać o rząd wielkości więcej danych do trenowania generatywnych modeli sztucznej inteligencji [88].

- Jensen Huang, założyciel i dyrektor generalny NVIDIA

Za pomocą Pandas można zarządzać i analizować zbiory danych znacznie wykraczające poza możliwości Excela. Podczas gdy Excel jest zwykle w stanie obsłużyć do 1 miliona wierszy danych, Pandas może z łatwością obsługiwać zestawy danych (Rys. 9.1-2, Rys. 9.1-10) zawierające dziesiątki milionów wierszy [89]. Ta zdolność pozwala użytkownikom na wykonywanie zaawansowanych analiz danych i wizualizacji na dużych zbiorach danych, zapewniając głęboki wgląd i ułatwiając podejmowanie decyzji w oparciu o dane. Ponadto Pandas ma silne wsparcie społeczności [90]: setki milionów programistów i analityków na całym świecie (Kaggle.com, Google Collab, Microsoft® Azure™ Notebooks, Amazon SageMaker) korzysta z niego online lub offline każdego dnia, zapewniając dużą liczbę gotowych rozwiązań dla każdego problemu biznesowego.

Sercem większości procesów analitycznych Pythona jest ustrukturyzowana forma danych zwana DataFrame, dostarczana przez bibliotekę Pandas. Jest to potężne i elastyczne narzędzie do organizowania, analizowania i wizualizacji danych tabelarycznych.

DataFrame: uniwersalny format danych tabelarycznych

DataFrame to centralna struktura w bibliotece Pandas, która jest dwuwymiarową tabelą (Rys. 3.4-6), w której wiersze odpowiadają poszczególnym obiektom lub rekordom, a kolumny odpowiadają ich cechom, parametrom lub kategoriom. Struktura ta wizualnie przypomina arkusze kalkulacyjne Excel, ale jest znacznie lepsza pod względem elastyczności, skalowalności i funkcjonalności.

DataFrame to sposób reprezentowania i przetwarzania danych tabelarycznych przechowywanych w pamięci RAM komputera.

DataFrame to sposób reprezentowania i przetwarzania danych tabelarycznych przechowywanych w pamięci RAM komputera. W tabeli wiersze mogą odzwierciedlać np. elementy projektu budowlanego, a kolumny - ich właściwości: kategorie, wymiary, współrzędne, koszt, terminy itd. Co więcej, taka tabela może zawierać zarówno informacje o jednym projekcie (rys. 4.1-13), jak i dane o milionach obiektów z tysiący różnych projektów (rys. 9.1-10). Dzięki wektoryzowanym operacjom Pandas łatwo jest filtrować, grupować i agregować takie ilości informacji z dużą prędkością.

	ID	Name	Category	Family Name	Height	BoundingBoxMin_X	BoundingBoxMin_Y	BoundingBoxMin_Z	Level
axis=1	431144	Single-Flush	OST_Doors	Single-Flush	6.88976378	20.1503	-10.438	9.84252	Level 1
	431198	Single-Flush	OST_Doors	Single-Flush	6.88976378	13.2281	-1.1207	9.84252	Level 2
	457479	Single Window	OST_Windows	Single Window	8.858267717	-11.434	-11.985	9.80971	Level 2
	485432	Single Window	OST_Windows	Single Window	8.858267717	-11.434	4.25986	9.80971	Level 1
	490150	Single-Flush	OST_Doors	Single-Flush	6.88976378	-1.15748	-2.9565	-1E-16	Level 1
	493697	Basic Wall	OST_Walls	Basic Wall	Basic Wall	-38.15	20.1656	-4.9213	Level 1
	497540	Basic Wall	OST_Walls	Basic Wall	Basic Wall	-4.5212	0.0708	9.84252	Level 1

Rys. 3.4-6 Construction Project as a DataFrame to dwuwymiarowa tabela z elementami w wierszach i atrybutami w kolumnach.

Nvidia szacuje, że już dziś do 30 procent wszystkich zasobów obliczeniowych jest wykorzystywanych do przetwarzania danych strukturalnych - ramek danych - i udział ten stale rośnie.

Przetwarzanie danych jest tym, co prawdopodobnie stanowi jedną trzecią obliczeń wykonywanych w każdej firmie na świecie. Przetwarzanie danych i dane większości firm znajdują się w DataFrame, w formacie tabeli

- Jensen Huang, CEO firmy Nvidia [91]

Wymień kilka kluczowych funkcji DataFrame w Pandas:

- **Kolumny:** w DataFrame dane są zorganizowane w kolumny, z których każda ma unikalną nazwę. Kolumny atrybutów mogą zawierać dane różnych typów, podobnie jak kolumny w bazach danych lub kolumny w tabelach.
- **Pandas Series** to jednowymiarowa struktura danych w Pandas, podobna do listy lub kolumny w tabeli, gdzie każda wartość odpowiada innemu indeksowi

Pandas Series ma ponad 400 atrybutów i metod, dzięki czemu praca z danymi jest niezwykle elastyczna. Możesz bezpośrednio zastosować jedną z czterystu dostępnych funkcji do kolumny, wykonywać operacje matematyczne, filtrować dane, zastępować wartości, pracować z datami, ciągami znaków i nie tylko. Ponadto Series obsługuje operacje wektorowe, co znacznie przyspiesza przetwarzanie dużych zestawów danych w porównaniu do obliczeń cyklicznych. Na przykład, można łatwo pomnożyć wszystkie wartości przez liczbę, zastąpić brakujące dane lub zastosować złożone transformacje bez pisania skomplikowanych pętli.
- **Wiersze:** w DataFrame mogą być indeksowane unikalnymi wartościami. Indeks ten umożliwia szybką zmianę i dostosowanie danych w określonych wierszach.
- **Indeks:** Domyślnie podczas tworzenia DataFrame Pandas przypisuje każdemu wierszowi indeks od 0 do N-1 (gdzie N to liczba wszystkich wierszy w DataFrame). Indeks można jednak zmienić, aby uwzględnić specjalne oznaczenia, takie jak daty lub unikalne cechy.
- **Indeksowanie wierszy** w DataFrame oznacza, że każdy wiersz ma przypisaną unikalną nazwę lub etykietę, która jest nazywana indeksem DataFrame.
- **Typy danych:** DataFrame obsługuje różne typy danych, w tym: `int`, `float`, `bool`, `datetime64` i `object` dla danych tekstowych. Każda kolumna DataFrame ma swój własny typ danych, który określa, jakie operacje można wykonać na jej zawartości.
- **Operacje na danych:** DataFrame obsługuje szeroki zakres operacji przetwarzania danych, w tym agregację (`groupby`), łączenie (`merge` i `join`), konkatenację (`concat`), dzielenie-zastosowanie-łączenie i wiele innych technik transformacji danych.
- **Manipulacja rozmiarem:** DataFrame umożliwia dodawanie i usuwanie kolumn oraz wierszy, dzięki czemu jest dynamiczną strukturą, którą można modyfikować zgodnie z potrzebami analizy danych.
- **Wizualizacja danych:** za pomocą wbudowanych technik wizualizacji lub interfejsu z popularnymi bibliotekami wizualizacji danych, takimi jak Matplotlib lub Seaborn, DataFrame można łatwo przekształcić w wykresy i diagramy w celu graficznej prezentacji danych.
- **Dane wejściowe i wyjściowe:** Pandas zapewnia funkcje odczytu importu i eksportu danych do różnych formatów plików, takich jak CSV, Excel, JSON, HTML i SQL, potencjalnie czyniąc DataFrame centralnym centrum gromadzenia i dystrybucji danych.

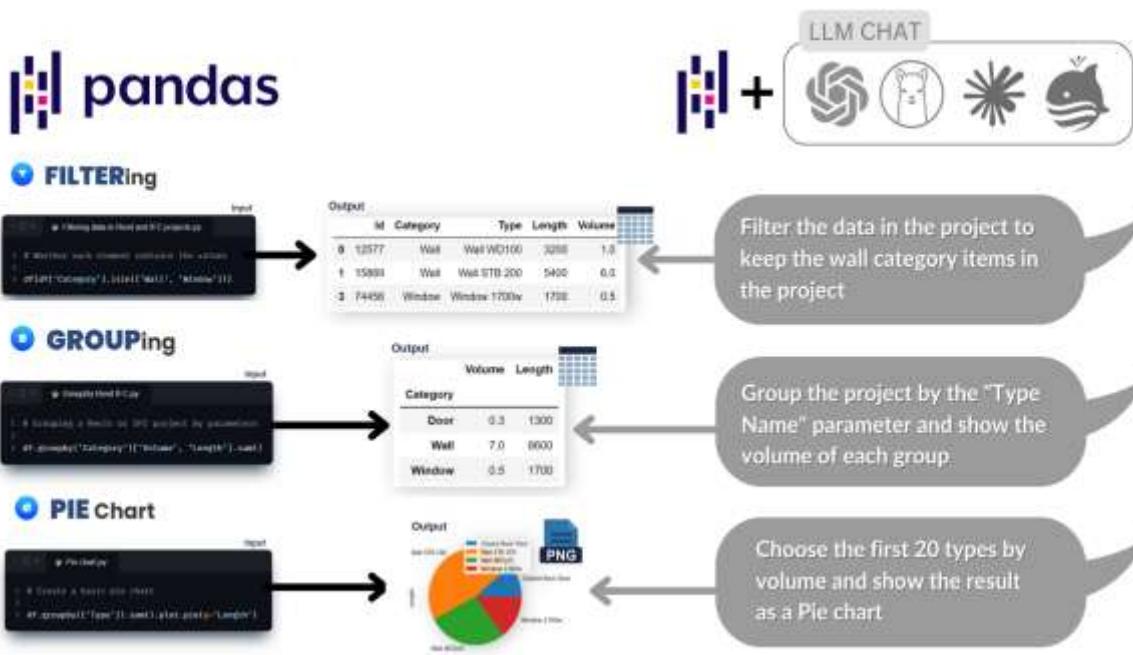
W przeciwieństwie do CSV i XLSX, Pandas DataFrame zapewnia większą elastyczność i wydajność podczas pracy z danymi: może obsługiwać duże ilości informacji w pamięci RAM, obsługuje rozszerzone typy danych (w tym daty, wartości logiczne i szeregi czasowe) oraz zapewnia szerokie możliwości filtrowania, agregowania, łączenia i wizualizacji danych. Podczas gdy CSV nie przechowuje informacji o typach i strukturze

danych, a XLSX jest często przeciążony formatowaniem i ma niską skalowalność, DataFrame pozostaje optymalnym wyborem do szybkiej analizy, automatyzacji procesów i integracji z modelami AI (Rys. 3.4-7). W kolejnych rozdziałach szczegółowo omówiony zostanie każdy z tych aspektów danych, a w części 8 książki szczegółowo omówione zostaną podobne formaty, takie jak Parquet, Apache Orc, JSON, Feather, HDF5 i hurtownie danych (rys. 8.1-2).

		XLSX	CSV	Pandas DataFrame
	Storage	Tabular	Tabular	Tabular
	Usage	Office tasks, data presentation	Simple data exchange	Data analysis, manipulation
	Compression	Built-in	None	None (in-memory)
	Performance	Low	Medium	High (memory dependent)
	Complexity	High (formatting, styles)	Low	Low
	Data Type Support	Limited	Very limited	Extended
	Scalability	Low	Low	Medium (memory limited)

Rys. 3.4-7 DataFrame to optymalny wybór do manipulacji danymi z wysoką wydajnością i zaawansowaną obsługą typów danych.

Dzięki swojej elastyczności, mocy i łatwości użycia, biblioteka Pandas i format DataFrame stały się de facto standardem w analizie danych w Pythonie. Są one idealne zarówno do tworzenia prostych raportów, jak i budowania złożonych potoków analitycznych, zwłaszcza w połączeniu z modelami LLM.



Rys. 3.4-8 LLM upraszczają interakcję z Pandas: zamiast kodu wystarczy zapytanie tekowe.

Obecnie Pandas jest aktywnie wykorzystywany w czatach opartych na LLM, takich jak ChatGPT, LLaMa, DeepSeek, QWEN i innych. W wielu przypadkach, gdy model otrzymuje zapytanie związane z przetwarzaniem tabel, walidacją danych lub analizą, generuje kod dokładnie przy użyciu biblioteki Pandas. To sprawia, że DataFrame jest naturalnym "językiem" do reprezentowania danych w dialogach ze sztuczną inteligencją (Rys. 3.4-8).

Nowoczesne technologie danych, takie jak Pandas, ułatwiają analizę, automatyzację i integrację danych z procesami biznesowymi. Szybko dostarczają wyniki, zmniejszają obciążenie specjalistów i zapewniają powtarzalność operacji.

Kolejne kroki: budowanie zrównoważonych ram danych

W tej części przyjrzaliśmy się kluczowym typom danych wykorzystywanych w branży budowlanej, poznaliśmy różne formaty ich przechowywania i przeanalizowaliśmy rolę nowoczesnych narzędzi, w tym LLM i IDE, w przetwarzaniu informacji. Dowiedzieliśmy się, że efektywne zarządzanie danymi jest podstawą świadomego podejmowania decyzji i automatyzacji procesów biznesowych. Organizacje, które są w stanie ustrukturyzować i zorganizować swoje dane, zyskują znaczącą przewagę konkurencyjną na etapach przetwarzania i transformacji danych.

Podsumowując tę część, warto podkreślić główne praktyczne kroki, które pomogą ci zastosować omówione podejścia w codziennych zadaniach:

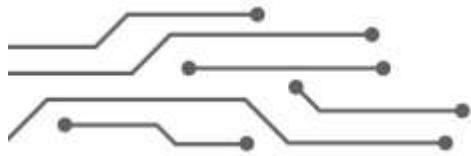
- Przeprowadzenie audytu danych procesów
- Sporządź spis wszystkich typów danych używanych w projektach.

- Określenie, które typy i modele danych są najbardziej krytyczne dla procesów biznesowych.
 - Identyfikacja obszarów problemowych, w których informacje są często nieuporządkowane, słabo ustrukturyzowane lub niedostępne.
- **Zacznij tworzyć strategię zarządzania danymi**
- Podnoszenie kwestii politycznych i standardów postępowania z różnymi typami danych
 - Przeanalizuj, które z przepływów pracy można usprawnić, przekształcając dane nieustrukturyzowane w dane ustrukturyzowane.
 - Stworzenie polityki przechowywania i dostępu do danych uwzględniającej bezpieczeństwo i poufność.
- **Zainstaluj i opanuj podstawowe narzędzia do pracy z danymi**
- Wybierz odpowiednie IDE, które odpowiada Twoim zadaniom (np. zainstaluj VS Code lub Jupyter Notebook).
 - Spróbuj założyć lokalną stronę LLM, aby poufnie przetwarzać swoje dane osobowe
 - Zacznij eksperymentować z biblioteką Pandas, aby przetwarzać dane tabelaryczne XLSX
 - Opisz LLM typowe zadania wykonywane w arkuszach kalkulacyjnych lub bazach danych i poproś LLM o zautomatyzowanie pracy przy użyciu Pandas.

Zastosowanie tych kroków pozwoli stopniowo zmienić podejście do pracy z danymi, przechodząc od rozproszonych, nieustrukturyzowanych zestawów danych do ujednoliconego ekosystemu, w którym dane stają się dostępny i zrozumiałym zasobem. Zacznij od małych kroków - stwórz swój pierwszy DataFrame w Pandas, uruchom lokalny LLM, zautomatyzuj swoje pierwsze rutynowe zadanie za pomocą Pythona (np. arkusze kalkulacyjne Excel).

Czwarta część książki skupi się na jakości danych, organizacji, strukturyzacji i modelowaniu danych. Skoncentrujemy się na metodologiach, które przekształcają różne źródła danych - od plików PDF i tekstów po obrazy i modele CAD - w ustrukturyzowane zestawy danych odpowiednie do analizy i automatyzacji. Zbadamy również, w jaki sposób wymagania dotyczące danych są sformalizowane, jak budowane są modele koncepcyjne i logiczne w projektach budowlanych oraz w jaki sposób nowoczesne modele językowe (LLM) mogą pomóc w tym procesie.





III CZĘŚĆ

JAKOŚĆ DANYCH: ORGANIZACJA, STRUKTURA, MODELOWANIE

Czwarta część koncentruje się na metodologiach i technologiach, które zapewniają przekształcanie rozproszonych informacji w ustrukturyzowane zbiory danych o wysokiej jakości. Szczegółowo omówiono procesy tworzenia i dokumentowania wymagań dotyczących danych jako podstawy efektywnej architektury informacji w projektach budowlanych. Przedstawiono praktyczne metody ekstrakcji ustrukturyzowanych informacji z różnych źródeł (PDF -dokumenty, obrazy, pliki tekstowe, CAD -modele) wraz z przykładami implementacji. Przeanalizowano wykorzystanie wyrażeń regularnych (RegEx) i innych narzędzi do automatycznej walidacji i weryfikacji danych. Opisano krok po kroku proces modelowania danych na poziomie koncepcyjnym, logicznym i fizycznym, z uwzględnieniem specyfiki branży budowlanej. Przedstawiono konkretne przykłady wykorzystania modeli językowych (LLM) do automatyzacji procesów strukturyzacji i walidacji informacji. Zaproponowano skuteczne podejście do wizualizacji wyników analiz, zwiększające dostępność informacji analitycznych dla wszystkich poziomów zarządzania projektami budowlanymi

ROZDZIAŁ 3.1.

KONWERSJA DANYCH DO POSTACI USTRUKTURYZOWANEJ

W erze gospodarki opartej na danych, dane stają się podstawą podejmowania decyzji, a nie przeszkodą. Zamiast stale dostosowywać informacje do każdego nowego systemu i jego formatów, firmy coraz częściej dążą do stworzenia jednego ustrukturyzowanego modelu danych, który służy jako uniwersalne źródło prawdy dla wszystkich procesów. Nowoczesne systemy informatyczne są projektowane nie wokół formatów i interfejsów, ale wokół znaczenia danych - ponieważ struktura może się zmieniać, ale znaczenie informacji pozostaje takie samo przez znacznie dłuższy czas.

Kluczem do efektywnej pracy z danymi nie jest konwertowanie i przekształcanie ich w nieskończoność, ale ich prawidłowa organizacja od samego początku: stworzenie uniwersalnej struktury zdolnej do zapewnienia przejrzystości, automatyzacji i integracji na wszystkich etapach cyklu życia projektu.

Tradycyjne podejście wymusza ręczne dostosowywanie przy każdej nowej implementacji platformy: migracja danych, zmiana nazw atrybutów, dostosowywanie formatów. Kroki te nie poprawiają jakości samych danych, a jedynie maskują problemy, tworząc błędne koło niekończących się transformacji. W rezultacie firmy stają się zależne od konkretnych rozwiązań programowych, a transformacja cyfrowa ulega spowolnieniu.

W kolejnych rozdziałach przyjrzymy się, jak prawidłowo ustrukturyzować dane, a następnie jak stworzyć uniwersalne modele, zminimalizować zależność od platformy i skupić się na tym, co najważniejsze - danych jako strategicznym zasobie, wokół którego budowane są zrównoważone procesy.

Nauka przekształcania dokumentów, plików PDF, obrazów i tekstów w ustrukturyzowane formaty

W projektach budowlanych zdecydowana większość informacji występuje w formie nieustrukturyzowanej: dokumenty techniczne, zestawienia prac, rysunki, specyfikacje, harmonogramy, protokoły. Ich różnorodność - zarówno pod względem formatu, jak i treści - utrudnia integrację i automatyzację.

Proces konwersji do formatów ustrukturyzowanych lub częściowo ustrukturyzowanych może się różnić w zależności od rodzaju danych wejściowych i pożądanych wyników przetwarzania.

Przekształcanie danych z postaci nieustrukturyzowanej do ustrukturyzowanej jest zarówno sztuką, jak i nauką. Proces ten różni się w zależności od rodzaju danych wejściowych i celu analizy i często zajmuje znaczną część pracy inżyniera danych (Rys. 3.2-5) i analityka, mając na celu stworzenie czystego, zorganizowanego zestawu danych.



Rys. 4.1-1 Konwersja nieuporządkowanego zeskanowanego dokumentu do ustrukturyzowanego formatu tabeli.

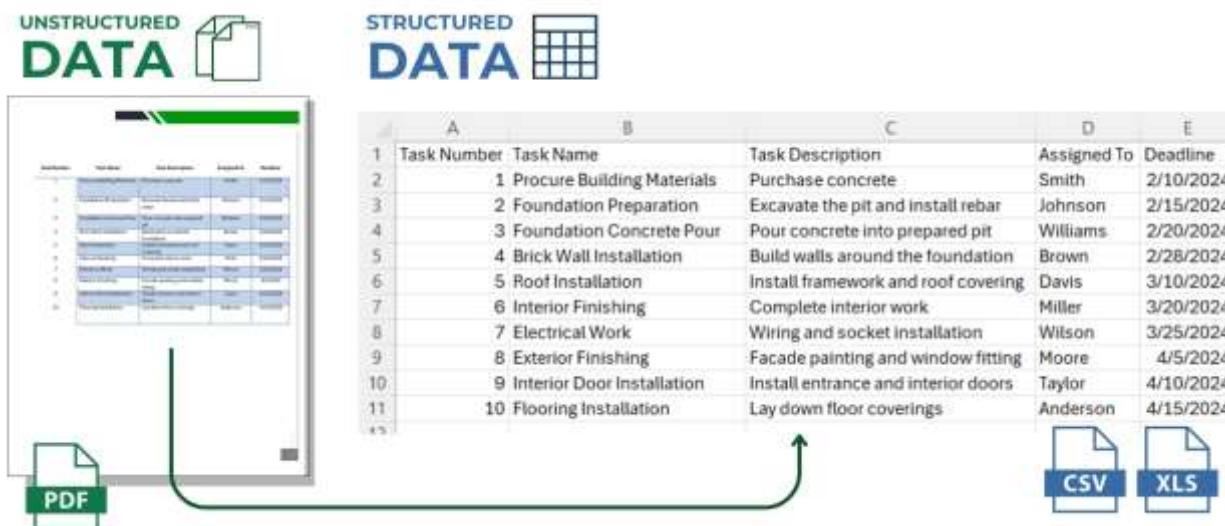
Przekształcanie dokumentów, plików PDF, obrazów i tekstów w ustrukturyzowany format (Rys. 4.1-1) jest procesem krok po kroku, który obejmuje następujące etapy:

- **Extract:** W tym kroku ładowany jest dokument źródłowy lub obraz zawierający nieuporządkowane dane. Może to być na przykład dokument PDF, zdjęcie, Rys. lub schemat.
- **Transformacja danych (Transform):** Następnie następuje etap przekształcania nieuporządkowanych danych w ustrukturyzowany format. Na przykład może to obejmować rozpoznawanie i interpretowanie tekstu z obrazów przy użyciu optycznego rozpoznawania znaków (OCR) lub innych metod przetwarzania.
- **Ładowanie i zapisywanie danych (Load):** ostatni krok obejmuje zapisanie przetworzonych danych w różnych formatach, takich jak CSV, XLSX, XML, JSON, do dalszej pracy, gdzie wybór formatu zależy od konkretnych wymagań i preferencji.

Proces ten, znany jako ETL (Extract, Transform, Load), odgrywa kluczową rolę w zautomatyzowanym przetwarzaniu danych i zostanie omówiony bardziej szczegółowo w rozdziale "ETL and Pipeline: Extract, Transform, Load". Następnie przyjrzymy się przykładom przekształcania dokumentów o różnych formatach w ustrukturyzowane dane.

Przykład konwersji dokumentu PDF na tabelę

Jednym z najczęstszych zadań w projektach budowlanych jest przetwarzanie specyfikacji w formacie PDF. Aby zademonstrować przejście od danych nieuporządkowanych do formatu ustrukturyzowanego, rozważmy praktyczny przykład: wyodrębnienie tabeli z dokumentu PDF i przekonwertowanie jej do formatu CSV lub Excel (rys. 4.1-2).



Rys. 4.1-2 W przeciwieństwie do formatu PDF, formaty CSV i XLSX są szeroko rozpowszechnione i łatwo integrują się z różnymi systemami zarządzania danymi.

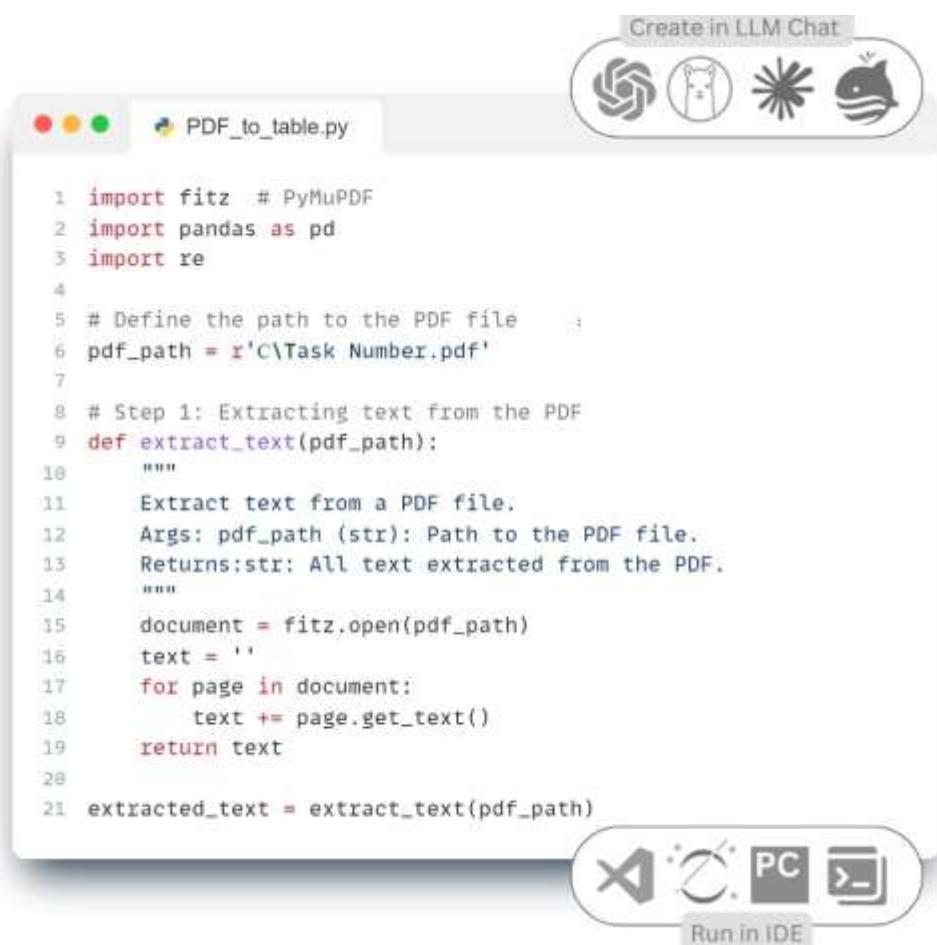
Modele językowe LLM, takie jak ChatGPT, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN znacznie upraszczają sposób, w jaki analitycy danych pracują z danymi, zmniejszając potrzebę głębokiego uczenia się języków programowania i umożliwiając rozwiązywanie wielu zadań za pomocą zapytań tekstowych.

Dlatego zamiast tracić czas na szukanie rozwiązań w Internecie (najczęściej na stronie StackOverflow lub forach i czatach tematycznych) lub kontaktować się ze specjalistami od przetwarzania danych, możemy skorzystać z możliwości nowoczesnych internetowych lub lokalnych LLM. Wystarczy zadać zapytanie, a model dostarczy gotowy kod do konwersji dokumentu PDF na format tabelaryczny.

- ❷ Wyślij następujące żądanie tekstowe do dowolnego modelu LLM (CHATGRT, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN lub dowolnego innego):

Proszę o napisanie kodu wyodrębniającego tekst z pliku PDF zawierającego tabelę. Kod powinien przyjmować ścieżkę do pliku jako argument i zwracać wyodrębnioną tabelę jako DataFrame ↴

- Odpowiedź modelu LLM w większości przypadków będzie miała postać kodu w języku Python , ponieważ język ten jest szeroko stosowany do przetwarzania danych, automatyzacji i pracy z różnymi formatami plików:



```

1 import fitz # PyMuPDF
2 import pandas as pd
3 import re
4
5 # Define the path to the PDF file
6 pdf_path = r'C\Task Number.pdf'
7
8 # Step 1: Extracting text from the PDF
9 def extract_text(pdf_path):
10     """
11     Extract text from a PDF file.
12     Args: pdf_path (str): Path to the PDF file.
13     Returns: str: All text extracted from the PDF.
14     """
15     document = fitz.open(pdf_path)
16     text = ''
17     for page in document:
18         text += page.get_text()
19     return text
20
21 extracted_text = extract_text(pdf_path)

```

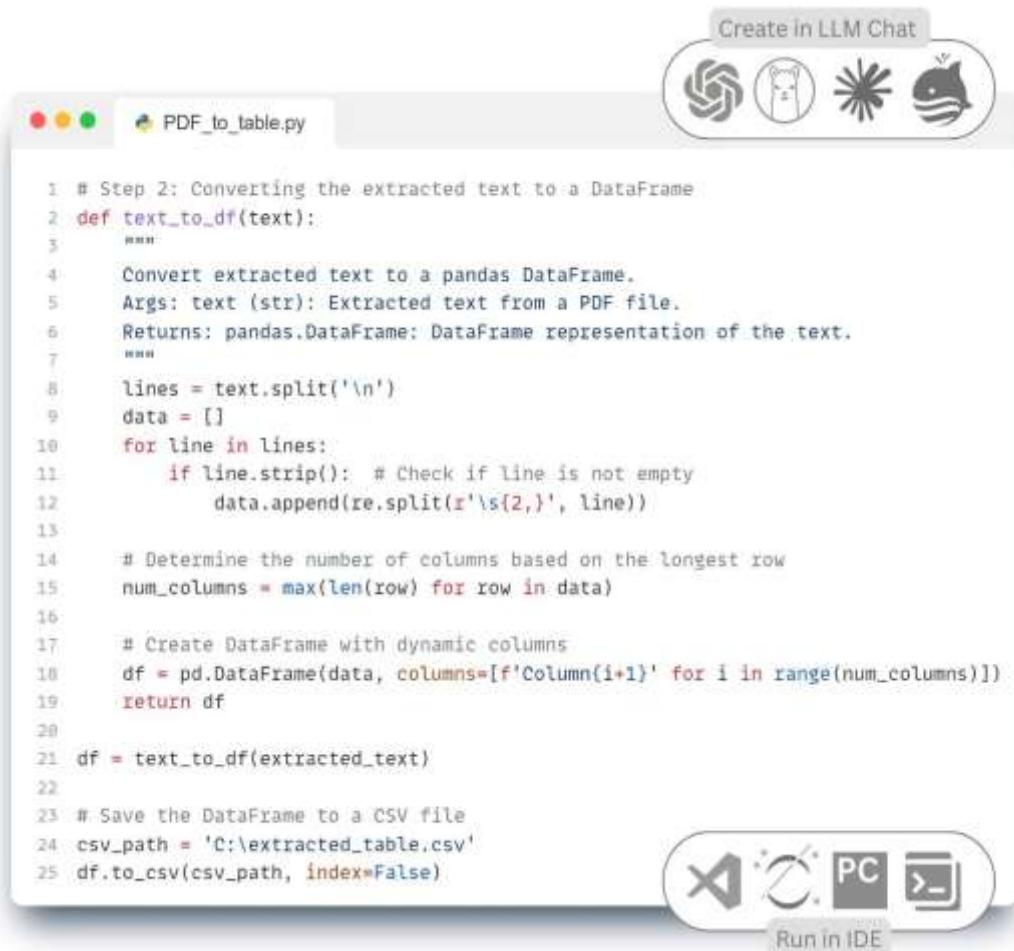
Rys. 4.1-3 Odpowiedź LLM w postaci kodu Pythona oraz jego bibliotek i pakietów (Pandas, Fitz) wyodrębnia tekst z pliku PDF .

Kod ten (Rys. 4.1-3) można uruchomić offline w jednym z popularnych IDE, o których wspomnialiśmy powyżej: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse z wtyczką PyDev, Thonny, Wing IDE, IntelliJ IDEA z wtyczką Python, JupyterLab lub popularnych narzędziach online: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- W kroku "Konwertuj" używamy popularnej biblioteki Pandas (któłą szczegółowo omówiliśmy w rozdziale "Python Pandas: niezastąpione narzędzie do pracy z danymi"), aby odczytać wyodrębniony tekst do DataFrame i zapisać DataFrame do pliku tabeli CSV lub XLXS:

Potrzebuję kodu, który przekonwertuje tabelę wynikową z pliku PDF do DataFrame. Dodaj również kod, aby zapisać DataFrame do pliku CSV. ↵

💡 Odpowiedź LLM :



The screenshot shows a code editor window titled "PDF_to_table.py". The code is as follows:

```

1 # Step 2: Converting the extracted text to a DataFrame
2 def text_to_df(text):
3     """
4         Convert extracted text to a pandas DataFrame.
5         Args: text (str): Extracted text from a PDF file.
6         Returns: pandas.DataFrame: DataFrame representation of the text.
7     """
8     lines = text.split('\n')
9     data = []
10    for line in lines:
11        if line.strip(): # Check if line is not empty
12            data.append(re.split(r'\s{2,}', line))
13
14    # Determine the number of columns based on the longest row
15    num_columns = max(len(row) for row in data)
16
17    # Create DataFrame with dynamic columns
18    df = pd.DataFrame(data, columns=[f'Column{i+1}' for i in range(num_columns)])
19    return df
20
21 df = text_to_df(extracted_text)
22
23 # Save the DataFrame to a CSV file
24 csv_path = 'C:\extracted_table.csv'
25 df.to_csv(csv_path, index=False)

```

At the top right of the window are icons for "Create in LLM Chat" and four AI models. At the bottom right are icons for "Run in IDE" and "PC".

Rys. 4.1-4 Konwersja wyodrębnionej tabeli z formatu PDF do formatu DataFrame i zapisanie tabeli w pliku CSV.

Jeśli podczas wykonywania kodu wystąpi błąd (rys. 4.1-3, rys. 4.1-4) - np. z powodu brakujących bibliotek lub nieprawidłowej ścieżki pliku - tekst błędu można po prostu skopiować wraz z kodem źródłowym i ponownie przesłać do modelu LLM. Model przeanalizuje komunikat o błędzie, wyjaśni na czym polega problem i zasugeruje poprawki lub dodatkowe kroki.

W ten sposób interakcja z AI LLM staje się kompletnym cyklem: →→test→ informacja zwrotna→ korekta - bez potrzeby posiadania głębokiej wiedzy technicznej.

Używając zwykłego zapytania tekstowego w czacie LLM i kilkunastu linii Pythona, które możemy uruchomić lokalnie w dowolnym IDE, przekonwertowaliśmy dokument PDF na tabelaryczny format CSV, który w przeciwieństwie do dokumentu PDF jest łatwy do odczytu maszynowego i szybko zintegrowany z dowolnym systemem zarządzania danymi.

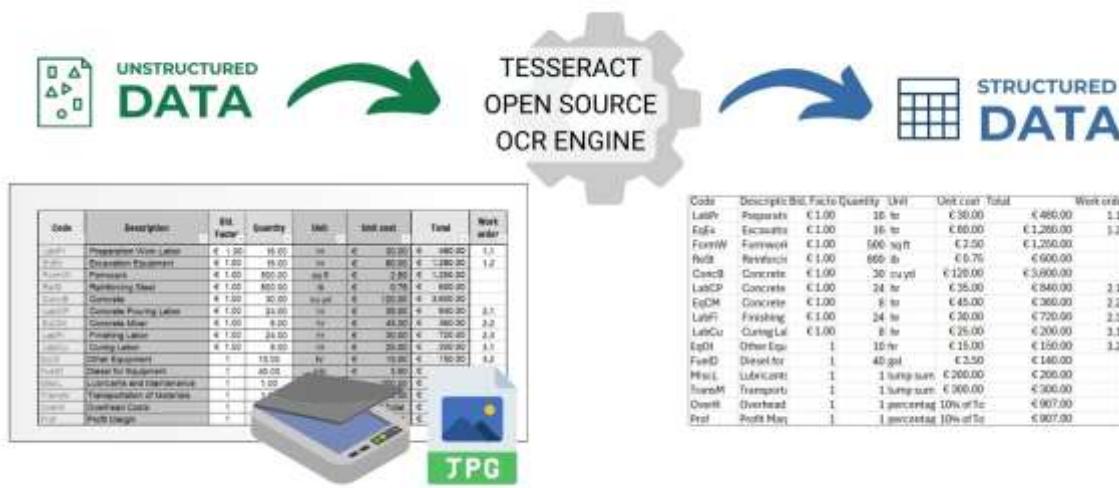
Możemy zastosować ten kod (Rys. 4.1-3, Rys. 4.1-4), kopiując go z dowolnego pokoju rozmów LLM, do dziesiątek lub tysięcy nowych dokumentów PDF na serwerze, automatyzując w ten sposób proces konwersji strumienia nieustrukturyzowanych dokumentów do ustrukturyzowanego formatu tabeli CSV.

Jednak dokumenty PDF nie zawsze zawierają tekst, częściej są to zeskanowane dokumenty, które muszą być przetwarzane jako obrazy. Chociaż obrazy są z natury nieustrukturyzowane, rozwój i zastosowanie bibliotek rozpoznawania pozwala nam wyodrębnić, przetwarzać i analizować ich zawartość, umożliwiając nam pełne wykorzystanie tych danych w naszych procesach biznesowych.

Konwersja obrazów JPEG, PNG do postaci strukturalnej

Obrazy są jedną z najpopularniejszych form nieustrukturyzowanych danych. W budownictwie i wielu innych branżach ogromna ilość informacji jest przechowywana w postaci zeskanowanych dokumentów, schematów, zdjęć i rysunków. Takie dane zawierają cenne informacje, ale nie mogą być bezpośrednio przetwarzane, np. w arkuszu kalkulacyjnym Excel lub bazie danych. Obrazy zawierają wiele złożonych informacji, ponieważ ich zawartość, kolory, tekstury są zróżnicowane i wymagane jest specjalne przetwarzanie w celu wyodrębnienia przydatnych informacji.

Trudność w wykorzystaniu obrazów jako źródła danych polega na braku struktury. Obrazy nie przekazują znaczenia w bezpośredni, łatwo wymierny sposób, który komputer może natychmiast zrozumieć lub przetworzyć, tak jak robi to arkusz kalkulacyjny Excel lub tabela bazy danych. Aby przekształcić nieustrukturyzowane dane obrazu w ustrukturyzowaną formę, należy użyć specjalnych bibliotek zdolnych do interpretacji zawartych w nich informacji wizualnych (Rys. 4.1-5).



Rys. 4.1-5 Konwersja zeskanowanych dokumentów i obrazów do formatów strukturalnych jest możliwa dzięki specjalnym narzędziom OCR.

Technologia OCR (Optical Character Recognition) służy do wyodrębniania tekstu z obrazów. Umożliwia ona

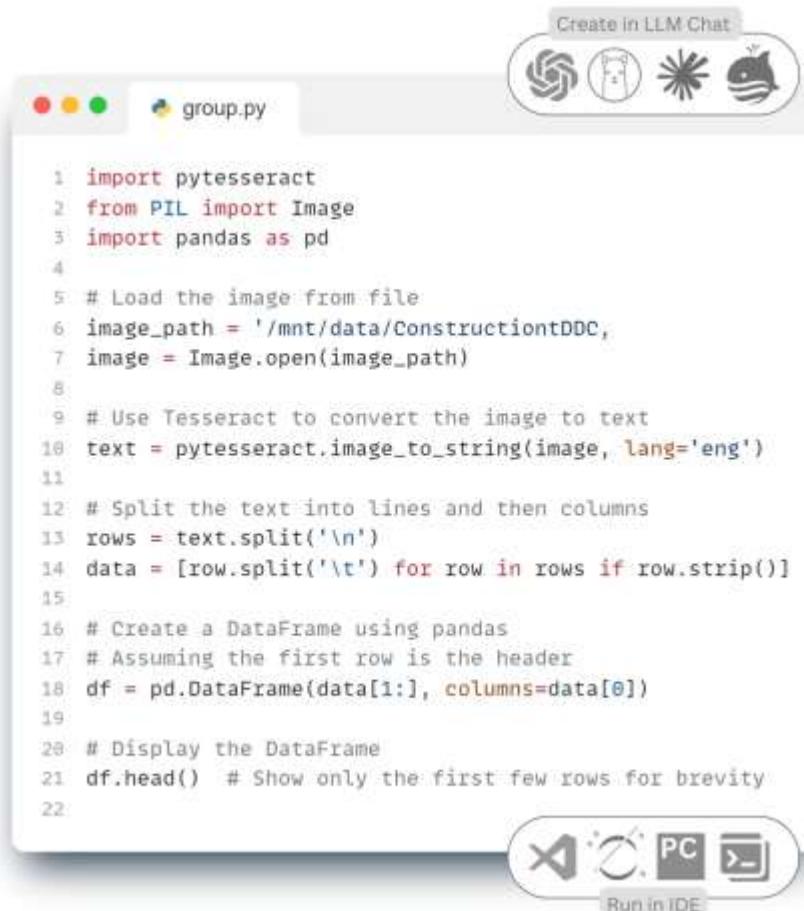
rozpoznawanie liter i cyfr w skanach dokumentów, zdjęciach i plikach PDF, przekształcając je w edytowalny i nadający się do odczytu maszynowego tekst. Technologie OCR są od dawna wykorzystywane w automatyzacji dokumentów, a dziś można je łatwo zintegrować z dowolnymi procesami biznesowymi i aplikacjami Python. Jednym z najpopularniejszych narzędzi OCR jest Tesseract, open source, pierwotnie opracowany przez HP™, a obecnie wspierany przez Google™. Obsługuje ono ponad 100 języków i charakteryzuje się wysoką dokładnością rozpoznawania.

Poprośmy czat LLM o napisanie przykładowego kodu do pobierania danych z zeskanowanej lub sfotografowanej tabeli w ustrukturyzowany sposób.

- 👉 Wyślij prośbę tekstową na czat LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN lub dowolny inny):

Napisz kod konwertujący obraz JPEG zawierający tabelę na tabelę DataFrame ↴

- 💡 Odpowiedź LLM w większości przypadków sugeruje użycie biblioteki Pytesseract do rozpoznawania tekstu na obrazach :



```

1 import pytesseract
2 from PIL import Image
3 import pandas as pd
4
5 # Load the image from file
6 image_path = '/mnt/data/ConstructionDDC,
7 image = Image.open(image_path)
8
9 # Use Tesseract to convert the image to text
10 text = pytesseract.image_to_string(image, lang='eng')
11
12 # Split the text into lines and then columns
13 rows = text.split('\n')
14 data = [row.split('\t') for row in rows if row.strip()]
15
16 # Create a DataFrame using pandas
17 # Assuming the first row is the header
18 df = pd.DataFrame(data[1:], columns=data[0])
19
20 # Display the DataFrame
21 df.head() # Show only the first few rows for brevity
22

```

Rys. 4.1-6 Konwertowanie tekstu wyodrębnionego z tabeli obrazów lub zdjęć na ustrukturyzowaną reprezentację tabelaryczną.

W tym przykładzie - kod (Rys. 4.1-6) pochodzący z LLM, wykorzystuje bibliotekę pytesseract (Tesseract for Python) do konwersji obrazu na tekst przy użyciu OCR (optyczne rozpoznawanie znaków) oraz bibliotekę Pandas do konwersji tego tekstu na ustrukturyzowaną formę, tj. DataFrame.

Proces konwersji zwykle obejmuje wstępne przetwarzanie w celu poprawy jakości obrazu, po czym stosowane są różne algorytmy wykrywania wzorców, ekstrakcji cech lub rozpoznawania obiektów. W rezultacie nieustrukturyzowane informacje wizualne są przekształcane w ustrukturyzowane dane.

Chociaż PDF i obrazy są kluczowymi źródłami nieustrukturyzowanych informacji, prawdziwym mistrzem pod względem objętości jest tekst generowany w wiadomościach e-mail, czatach, spotkaniach, komunikatorach. Dane te są nie tylko liczne - są rozproszone, nieformalne i wyjątkowo słabo ustrukturyzowane.

Konwersja danych tekstowych do postaci ustrukturyzowanej

Oprócz dokumentów PDF z tabelami (rys. 4.1-2) i zeskanowanych wersji formularzy tabelarycznych (rys. 4.1-5), znaczna część informacji w dokumentacji projektowej jest prezentowana w formie tekstu. Mogą to być zarówno spójne zdania w dokumentach tekstowych, jak i fragmentaryczne zapisy rozproszone na rysunkach i schematach. W nowoczesnych warunkach przetwarzania danych jednym z najczęstszych zadań jest przekształcenie takiego tekstu w ustrukturyzowany format odpowiedni do analizy, wizualizacji i podejmowania decyzji.

Centralnym elementem tego procesu jest taksonomia, system klasyfikacji, który pozwala na organizowanie informacji w kategorie i podkategorie w oparciu o wspólne cechy.

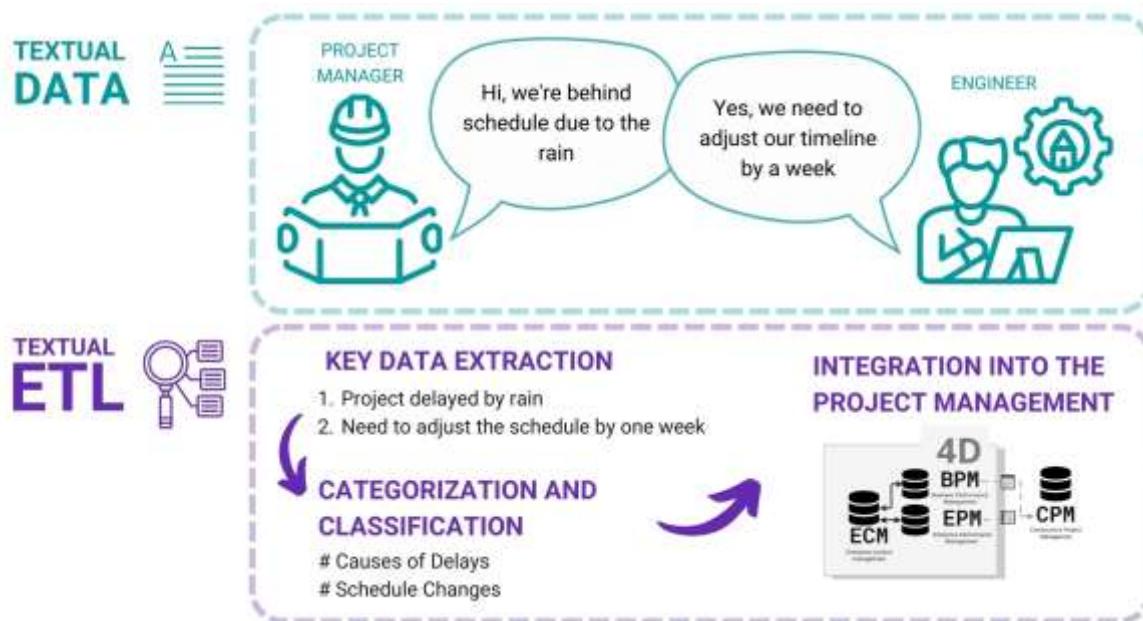
Taksonomia to hierarchiczna struktura klasyfikacyjna wykorzystywana do grupowania i organizowania obiektów. W kontekście przetwarzania tekstu służy jako podstawa do systematycznego kategoryzowania elementów w kategorie semantyczne, upraszczając w ten sposób analizę i poprawiając jakość przetwarzania danych.

Tworzeniu taksonomii towarzyszą etapy ekstrakcji encji, kategoryzacji i kontekstualizacji. Aby zamodelować proces wyodrębniania informacji z danych tekstowych, poniższe kroki są podobne do tych, które zastosowaliśmy już do strukturyzacji danych z dokumentów PDF:

- **Extract:** należy przeanalizować dane tekstowe, aby wyodrębnić informacje o opóźnieniach i zmianach w harmonogramie projektu.
- **Kategoryzacja i klasyfikacja (Transofrm):** kategoryzacja otrzymanych informacji, np. przyczyny opóźnień i zmian w harmonogramie.
- **Integracja (Load):** na koniec przygotowujemy ustrukturyzowane dane do integracji z zewnętrznymi systemami zarządzania danymi.

Rozważmy sytuację: mamy dialog pomiędzy kierownikiem projektu i inżynierem omawiający problemy z harmonogramem. Naszym celem jest wyodrębnienie kluczowych elementów (przyczyny opóźnień, korekty harmonogramu) i przedstawienie ich w uporządkowany sposób (Rys. 4.1-7).

Wykonajmy ekstrakcję w oparciu o oczekiwane słowa kluczowe, utwórzmy DataFrame, aby zasymulować ekstrakcję danych, a po transformacji nową tabelę DataFrame, która będzie zawierać kolumny daty, zdarzenia (np. powód opóźnienia) i działania (np. zmiana harmonogramu).



Rys. 4.1-7 Podkreślenie kluczowych informacji z tekstu o potrzebie dostosowania harmonogramów i zintegrowania zmian z systemem zarządzania projektem.

Oto kod rozwiązujejący problem przy użyciu zapytania tekstowego w jednym z modeli językowych, tak jak w poprzednich przykładach.

👉 Wyślij prośbę tekstową do dowolnego pokoju czatu LLM:

Mam rozmowę między kierownikiem: "Witam, mamy opóźnienie z powodu deszczu" i inżynierem: "Tak, musimy dostosować termin o tydzień". Potrzebuję skryptu, który przeanalizuje przyszłe podobne dialogi tekstowe, wyodrębnia z nich przyczyny opóźnień i niezbędnych korekt terminów, a następnie wygeneruje DataFrame z tych danych. DataFrame powinien następnie zostać zapisany do pliku CSV. ↵

- 2 Odpowiedź z LLM będzie zazwyczaj zawierać kod Pythona wykorzystujący wyrażenia regularne (re - Regex) i bibliotekę Pandas (pd) :

TEXTUAL DATA

A vertical stack of lines.

↓

Python Code:

```

import pandas as pd
import re

# Sample text data from a discussion between a manager and an engineer.
text_data = [
    "Hi, we're behind schedule due to the rain",
    "Yes, We need to adjust our timeline by a week"
]

# Extract key information elements
def extract_key_elements(text):
    delay_pattern = r"(behind schedule|delayed) due to the (\w+)"
    time_adjust_pattern = r"adjust our timeline by a (\w+)"

    delay_search = re.search(delay_pattern, text)
    time_adjust_search = re.search(time_adjust_pattern, text)

    delay_reason = delay_search.group(2) if delay_search else None
    time_adjustment = time_adjust_search.group(1)
        if time_adjust_search else None

    return delay_reason, time_adjustment

# Applying the function to our text data
extracted_data = [extract_key_elements(text) for text in text_data]

# Convert the data into a DataFrame
df_extracted = pd.DataFrame(extracted_data,
columns=['Cause of Delay', 'Time Adjustment'])

```

↓

Data Grid:

Create in LLM Chat

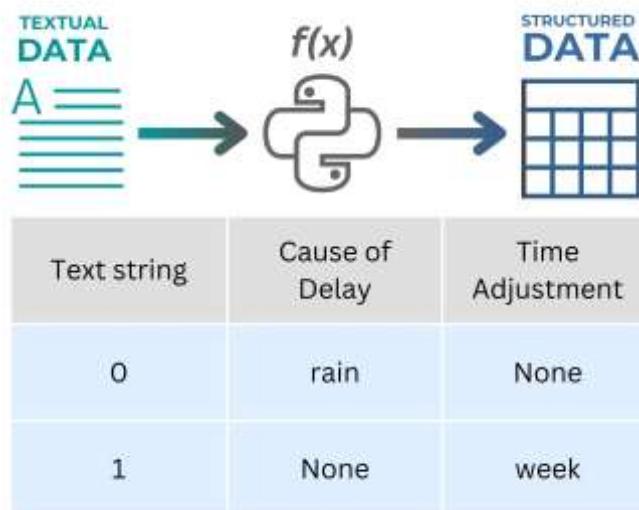
Run in IDE

Rys. 4.1-8 Podkreślenie kluczowych informacji z tekstu na temat konieczności dostosowania terminów w tabeli.

W tym przykładzie (Rys. 4.1-7) dane tekstowe zawierające korespondencję między kierownikiem projektu a inżynierem są analizowane w celu zidentyfikowania i wyodrębnienia konkretnych informacji, które mogą mieć wpływ na zarządzanie przyszłymi projektami z podobnymi dialogami. Korzystając z wyrażeń regularnych (więcej na temat wyrażeń regularnych omówimy w rozdziale "Wymagania strukturalne i wyrażenia regularne RegEx"), przyczyny opóźnień projektu i niezbędne korekty harmonogramu są identyfikowane za pomocą wzorców. Funkcja napisana w tym przykładzie wyodrębnia przyczynę opóźnienia lub dostosowanie czasu z ciągów na podstawie wzorców: wybierając słowo po "z powodu" jako przyczynę opóźnienia lub słowo po "przez" jako dostosowanie czasu.

Jeśli wiersz wspomina o opóźnieniu z powodu pogody, "deszcz" jest identyfikowany jako przyczyna; jeśli

wiersz wspomina o korekcie harmonogramu dla określonego okresu, okres ten jest wyodrębniany jako korekta czasu (Rys. 4.1-9). Brak któregokolwiek z tych słów w wierszu skutkuje wartością "Brak" dla odpowiedniej kolumny atrybutu.



Rys. 4.1-9 Tabela podsumowująca uzyskana jako DataFrame, po wykonaniu kodu, zawiera informacje o istnieniu opóźnień i koniecznych korektach czasu.

Strukturyzacja i parametryzacja warunków z tekstu (dialogu, listu, dokumentu) pozwala szybko wyeliminować opóźnienia w budowie: na przykład brak pracowników może wpływać na tempo prac przy złej pogodzie, więc firmy, znając parametry opóźnienia z dialogów (rys. 4.1-9) między brygadistą na placu budowy a kierownikiem projektu - z wyprzedzeniem mogą wzmocnić załogę w przypadku niekorzystnej prognozy.

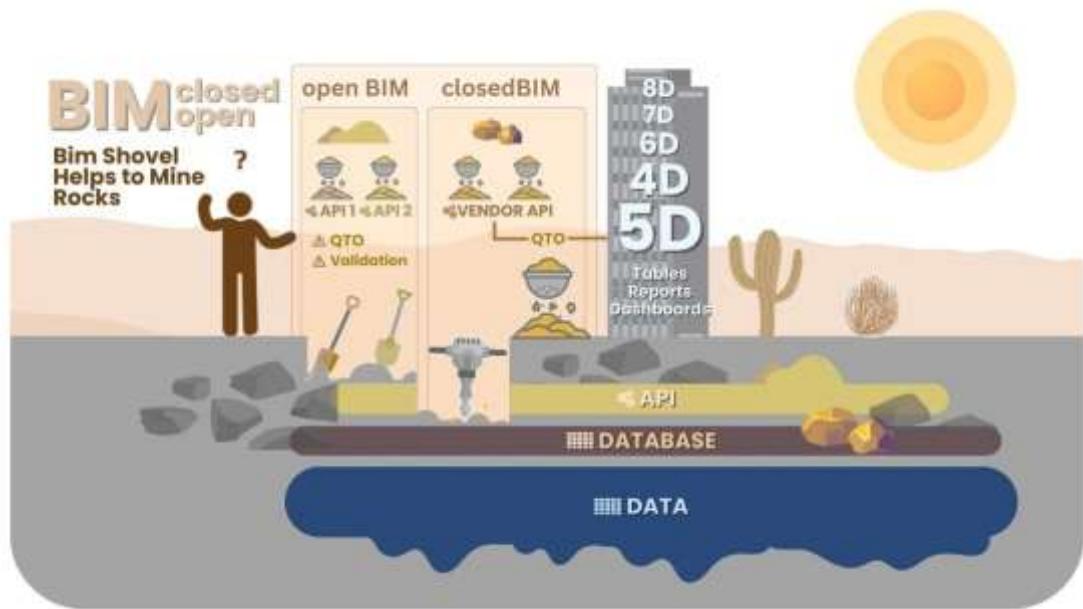
Konwersję dokumentów i obrazów do formatu strukturalnego można osiągnąć za pomocą stosunkowo prostych, otwartych i bezpłatnych narzędzi opartych na kategoryzacji.

Kategoryzacja elementów jest również kluczową częścią pracy z danymi projektowymi, zwłaszcza w kontekście korzystania z oprogramowania CAD (BIM).

Konwersja danych CAD (BIM) do postaci ustrukturyzowanej

Strukturyzacja i kategoryzacja danych CAD (BIM) jest bardziej złożonym zadaniem, ponieważ dane przechowywane w bazach danych CAD (BIM) są prawie zawsze w zamkniętych lub złożonych formatach parametrycznych, często łącząc elementy danych geometrycznych (częściowo ustrukturyzowane) i elementy metainformacji (dane częściowo ustrukturyzowane lub ustrukturyzowane) w tym samym czasie.

Natywne formaty danych w systemach CAD (BIM) są zwykle chronione i niedostępne do bezpośredniego użytku, chyba że specjalistyczne oprogramowanie lub API - interfejsy samego programisty (rys. 4.1-10). Taka izolacja danych tworzy zamknięte silosy pamięci masowej, które ograniczają swobodną wymianę informacji i hamują tworzenie kompleksowych procesów cyfrowych w firmie.



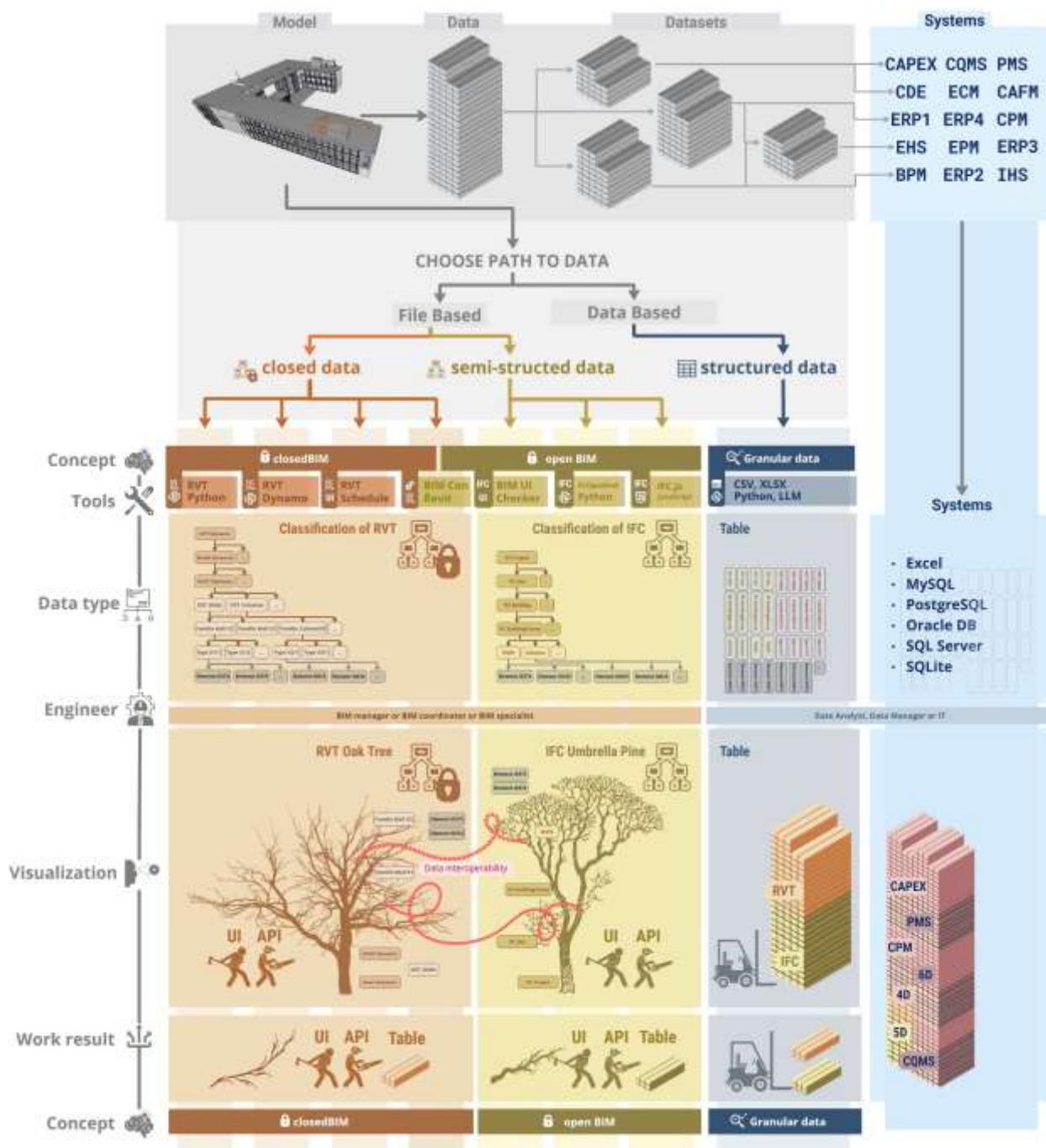
Rys. 4.1-10 Specjaliści CAD (BIM) mogą uzyskać dostęp do natywnych danych za pośrednictwem interfejsów API lub narzędzi dostawców.

W specjalnych formatach CAD (BIM) informacje o cechach i atrybutach elementów projektu są gromadzone w hierarchicznym systemie klasyfikacji, w którym jednostki o odpowiednich właściwościach znajdują się, podobnie jak owoce drzewa owocowego, w najnowszych węzłach gałęzi klasyfikacji danych (rys. 4.1-11).

Wyodrębnianie danych z takich hierarchii jest możliwe na dwa sposoby: albo ręcznie, klikając każdy węzeł, tak jakbyśmy przetwarzali drzewo, wycinając wybrane gałęzie kategorii i typów za pomocą siekiery. Alternatywnie, wykorzystanie interfejsów programowania aplikacji (API) zapewnia bardziej wydajne, zautomatyzowane podejście do pobierania i grupowania danych, ostatecznie przekształcając je w ustrukturyzowaną tabelę do wykorzystania w innych systemach.

Różne narzędzia, takie jak Dynamo, pyRvt, Pandamo (Pandas + Dynamo), ACC lub rozwiązań open source, takie jak IfcOpSh lub IFCjs dla formatu IFC, mogą być używane do wyodrębniania ustrukturyzowanych tabel danych z projektów CAD (BIM).

Nowoczesne narzędzia do eksportu i konwersji danych pozwalają uprościć przetwarzanie i przygotowanie danych poprzez podział zawartości modeli CAD na dwa kluczowe komponenty: informacje o geometrii i dane atrybutowe (rys. 4.1-13) - metainformacje opisujące właściwości elementów projektu (rys. 3.1-16). Te dwie warstwy danych pozostają powiązane za pomocą unikalnych identyfikatorów, dzięki czemu możliwe jest precyzyjne mapowanie każdego elementu z opisem geometrii (za pomocą parametrów lub wielokątów) do jego atrybutów: nazwy, materiału, etapu realizacji, kosztu itp. Takie podejście zapewnia integralność modelu i umożliwia elastyczne wykorzystanie danych zarówno do wizualizacji (dane modelu geometrycznego), jak i do zadań analitycznych lub zarządzania (ustrukturyzowane lub luźno ustrukturyzowane), pracując z dwoma typami danych oddzielnie lub równolegle.



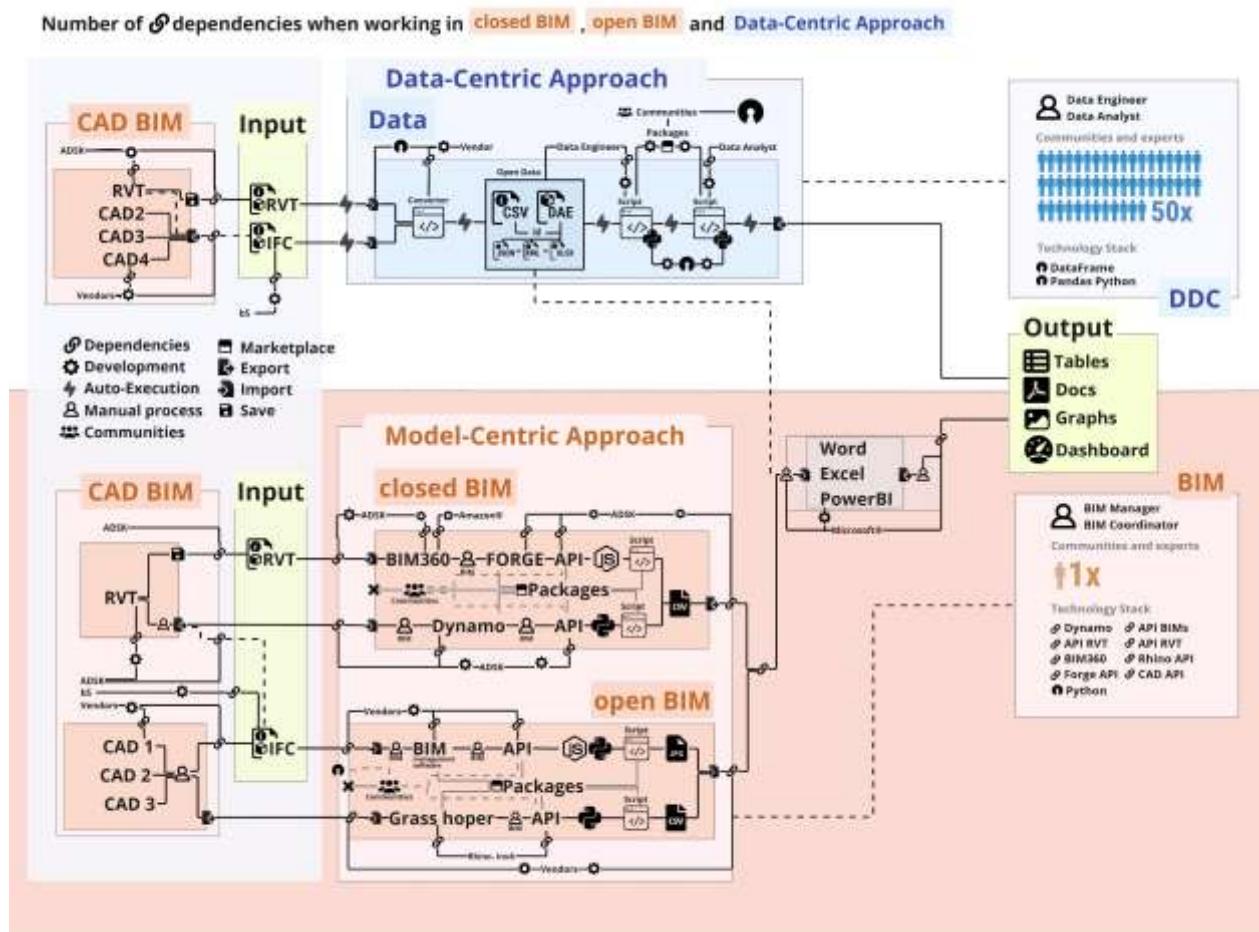
Rys. 4.1-11 Widok informacji z baz danych CAD (BIM) jest prezentowany użytkownikowi w postaci drzew klasyfikacyjnych.

Wraz z rozwojem technologii inżynierii odwrotnej i pojawieniem się zestawów SDK (Software Development Kit) do konwersji danych CAD - dostępność i konwersja danych z zamkniętych formatów oprogramowania CAD (BIM) stała się znacznie łatwiejsza. Obecnie możliwe jest legalne i bezpieczne konwertowanie danych z zamkniętych formatów do uniwersalnych formatów nadających się do analizy i wykorzystania w innych systemach. Historia pierwszych narzędzi inżynierii odwrotnej ("Open DWG") i walka o dominację nad

formatami producentów CAD została omówiona w rozdziale "Dane strukturalne: podstawa transformacji cyfrowej".

Narzędzia do inżynierii odwrotnej umożliwiają legalne pobieranie danych z zamkniętych, zastrzeżonych formatów, dzieląc informacje z mieszanego formatu CAD (BIM) na typy i formaty danych wymagane przez użytkownika, ułatwiając ich przetwarzanie i analizę.

Korzystając z inżynierii odwrotnej i bezpośredniego dostępu do informacji z baz danych CAD udostępnia informacje, umożliwiając otwarte dane i otwarte narzędzia, a także analizę danych przy użyciu standardowych narzędzi, tworzenie raportów, wizualizacje i integrację z innymi systemami cyfrowymi (rys. 4.1-12).



Rys. 4.1-12 Bezpośredni dostęp do danych CAD minimalizuje zależność od platform oprogramowania i zmierza w kierunku podejścia skoncentrowanego na danych.

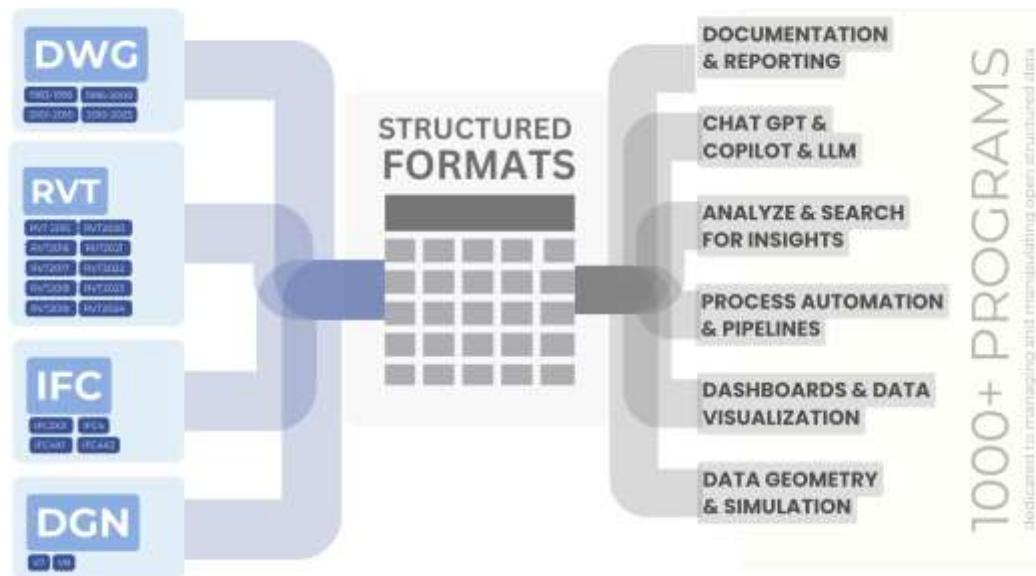
Od 1996 r. dla formatu DWG, od 2008 r. dla formatu DGN i od 2018 r. dla formatu RVT możliwa jest wygodna i wydajna konwersja początkowo zamkniętych formatów danych CAD na dowolne inne formaty, w tym formaty strukturalne, za pomocą narzędzi inżynierii odwrotnej (rys. 4.1-13). Obecnie prawie wszystkie główne firmy CAD (BIM) i duże firmy inżynierijne na świecie używają SDK - narzędzi inżynierii odwrotnej do wyodrębniania danych z zamkniętych formatów dostawców CAD (BIM) [92].



Rys. 4.1-13 Zastosowanie narzędzi inżynierii odwrotnej pozwala na konwersję baz danych programów CAD (BIM) do dowolnego wygodnego modelu danych.

Konwersja danych z zamkniętych, zastrzeżonych formatów do formatów otwartych i rozdzielenie mieszanych formatów CAD (BIM) na dane geometryczne i atrybuty metainformacji upraszcza proces pracy z nimi, udostępniając je do analizy, manipulacji i integracji z innymi systemami (rys. 4.1-14).

W dzisiejszej pracy z danymi CAD (BIM) osiągnęliśmy punkt, w którym nie musimy ubiegać się o pozwolenie od dostawców CAD (BIM), aby uzyskać dostęp do informacji z formatów CAD.



Rys. 4.1-14 Nowoczesne narzędzia SDK umożliwiają legalną konwersję danych z zastrzeżonych formatów baz danych CAD (BIM).

Obecne trendy w przetwarzaniu danych projektowych CAD są nadal kształtowane przez kluczowych graczy rynkowych - CAD - dostawców, którzy pracują nad wzmocnieniem swojej pozycji w świecie danych i tworzą nowe formaty i koncepcje.

Dostawcy rozwiązań CAD zmierzają w kierunku ustrukturyzowanych danych

Od 2024 r. branża projektowa i budowlana przechodzi znaczącą zmianę technologiczną w zakresie wykorzystania i przetwarzania danych. Zamiast swobodnego dostępu do danych projektowych, producenci systemów CAD koncentrują się na promowaniu kolejnych nowych koncepcji. Podejścia takie jak BIM (stworzony w 2002 roku) i otwarty BIM (stworzony w 2012 roku) stopniowo ustępują miejsca nowoczesnym rozwiązaniom technologicznym, które zaczynają promować dostawcy systemów CAD [93]:

- Przejście w kierunku wykorzystania "granularnych" danych, które pozwalają na efektywne zarządzanie informacjami i przejście w kierunku analizy danych.
- Pojawienie się formatu USD i wdrożenie podejścia Entity-component-system (ECS) dla elastycznej organizacji danych
- Aktywne wykorzystanie sztucznej inteligencji w przetwarzaniu danych, automatyzacji procesów i analizie danych.
- Rozwijanie interoperacyjności - lepsza interakcja między różnymi programami, systemami i bazami danych.

Każdy z tych aspektów zostanie omówiony bardziej szczegółowo w szóstej części książki "CAD i BIM: marketing, rzeczywistość i przyszłość danych projektowych w budownictwie". W tym rozdziale jedynie pokrótko nakreślmy ogólny wektor zmian: główni producenci CAD starają się obecnie przemyśleć sposób, w jaki informacje projektowe są ustrukturyzowane. Jedną z kluczowych zmian jest porzucenie klasycznego modelu przechowywania opartego na plikach na rzecz granularnej, zorientowanej analitycznie architektury danych, która zapewnia ciągły dostęp do poszczególnych komponentów modelu [93].

Istotą tego, co się dzieje, jest to, że branża wycofuje uciążliwe, specjalistyczne i parametryczne formaty, które wymagają geometrycznych jąder, na rzecz bardziej uniwersalnych, czytelnych dla maszyn i elastycznych rozwiązań.

Jednym z takich czynników zmian jest format USD (Universal Scene Description), pierwotnie opracowany w branży grafiki komputerowej, ale już uznany w zastosowaniach inżynierijnych dzięki rozwojowi platformy NVIDIA Omniverse (i Isaac Sim) do symulacji i wizualizacji [93]. W przeciwieństwie do parametrycznego IFC, USD oferuje prostszą strukturę i umożliwia opisywanie geometrii i właściwości obiektów w formacie JSON (Rys. 4.1-15), co ułatwia przetwarzanie informacji i przyspiesza ich integrację z procesami cyfrowymi. Nowy format umożliwia przechowywanie geometrii (oprócz BREP -NURBS - więcej szczegółów w części 6 książki) w postaci wielokątów MESH, a właściwości obiektów w JSON, co czyni go wygodniejszym dla zautomatyzowanych procesów i pracy w ekosystemach chmurowych [94].

Niektórzy producenci CAD i ERP używają już podobnych formatów (np. NWD, SVF, CP2, CPIXML), ale większość z nich pozostaje zamknięta i niedostępna do użytku zewnętrznego, co ogranicza możliwości integracji i ponownego wykorzystania danych. W tym kontekście USD może odgrywać taką samą rolę jak DXF w swoim czasie - otwartą alternatywę dla zastrzeżonych formatów, takich jak DWG.

General Information				Comparison / Notes
Year of format creation	1991	2016		IFC focuses on construction data, USD on 3D graphics
Creator-developer	TU Munich	Pixar		IFC was founded in Germany, USD in America
Prototypes and predecessors	IGES, STEP	PTEX, DAE, GLTF		IFC evolved from IGES/STEP, USD from PTEX/DAE/GLTF
Initiator in Construction	ADSK	ADSK		ADSK initiated the adoption of both formats in construction
Organizer of the Alliance	ADSK	ADSK		ADSK organized both alliances
Name of the Alliance	b5 (IAI)	AOUSD		Different alliances for each format
Year of Alliance Formation	1994	2023		The IFC alliance was formed in 1994, AOUSD for USD in 2023
Promoting in the construction	ADSK and Co	ADSK and Co		ADSK and Co actively promotes both formats in b5 (IAI) since the introduction

Purpose and Usage				Comparison / Notes
Purpose	Semantic description and interoperability	Data simplification, visualization unification		IFC for semantics and exchange; USD for simplification and visualization
Goals and Objectives	Interoperability and semantics	Unification for visualization and data processing		IFC focuses on semantics; USD on visualization
Use in Other Industries	Predominantly in construction	In film, games, VR/AR, and now in construction		USD is versatile and used in various fields
Supported Data Types	Geometry, object attributes, metadata	Geometry, shaders, animation, light, and camera		USD supports a wider range of data types suitable for complex visualizations; IFC focuses on construction-specific data

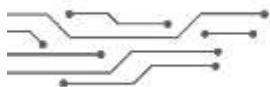
Rys. 4.1-15 Format USD jako próba zaspokojenia przez producentów CAD zapotrzebowania na interoperacyjność i niezależność danych projektowych od jąder geometrycznych.

Przejście głównych deweloperów na otwarte i uproszczone formaty USD, GLTF, OBJ, XML (zamknięte NWD, CP2, SVF, SVF2, CPIXML) i podobne formaty (rys. 3.1-17) odzwierciedla globalny trend i zapotrzebowanie branży na uproszczenie danych i zwiększenie ich dostępności. W nadchodzących latach możemy spodziewać się stopniowego odchodzenia od złożonych standardów parametrycznych i formatów zależnych od geometrycznych jąder na rzecz lżejszych i bardziej ustrukturyzowanych rozwiązań. Przejście to przyspieszy cyfryzację branży budowlanej, ułatwi automatyzację procesów i uprości wymianę danych.

Pomimo strategicznych planów dostawców CAD promowania nowych otwartych formatów, profesjonalisci z branżą budowlanej mogą również uzyskać pełny dostęp do danych z zamkniętych systemów CAD, bez potrzeby korzystania z narzędzi CAD (BIM), korzystając z narzędzi inżynierii odwrotnej.

Wszystkie te trendy nieuchronnie prowadzą do odejścia od nieporęcznych, monolitycznych modeli 3D na rzecz uniwersalnych, ustrukturyzowanych danych i wykorzystania formatów, które od dawna sprawdzają się w innych branżach. Gdy zespoły projektowe zaczynają postrzegać modele CAD nie tylko jako obiekty wizualne lub zestaw plików, ale jako bazy danych zawierające wiedzę i informacje, podejście do projektowania i zarządzania zmienia się diametralnie.

Gdy zespoły nauczą się wyodrębniać ustrukturyzowane dane z dokumentów, tekstów, rysunków i modeli CAD oraz uzyskają dostęp do baz danych, kolejnym kluczowym krokiem jest modelowanie danych i zapewnienie jakości. To właśnie ten etap w dużej mierze decyduje o szybkości przetwarzania i przekształcania informacji, które ostatecznie zostaną wykorzystane do podejmowania decyzji w konkretnych zadaniach aplikacji.



ROZDZIAŁ 3.2.

KLASYFIKACJA I INTEGRACJA: WSPÓLNY JĘZYK DLA DANYCH BUDOWLANYCH

Szybkość podejmowania decyzji zależy od jakości danych

Dzisiejsza architektura danych projektowych przechodzi fundamentalne zmiany. Branża odchodzi od nieporęcznych, odizolowanych modeli i zamkniętych formatów na rzecz bardziej elastycznych, nadających się do odczytu maszynowego struktur skoncentrowanych na analityce, integracji i automatyzacji procesów. Jednak samo przejście na nowe formaty nie gwarantuje wydajności - jakość samych danych jest nieuchronnie w centrum uwagi.

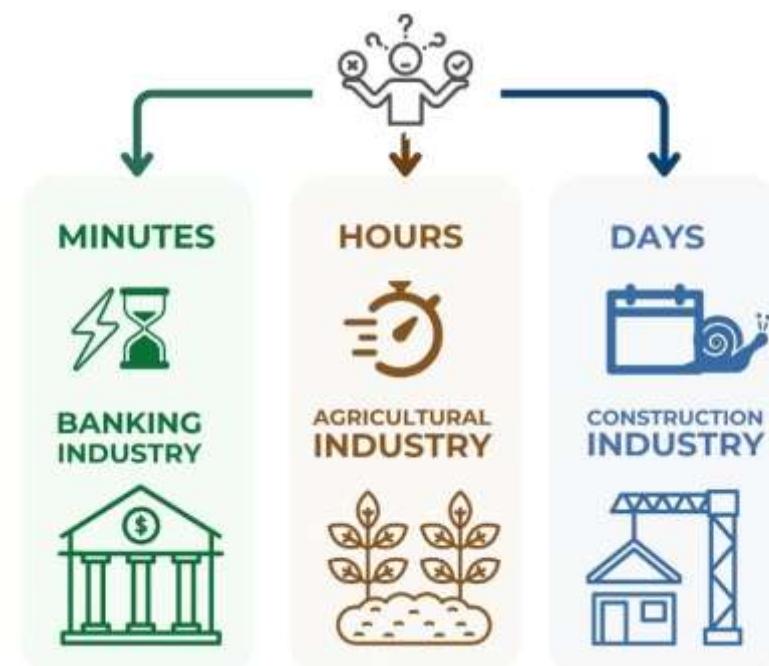
Na stronach tej książki dużo mówimy o formatach, systemach i procesach. Jednak wszystkie te wysiłki są bez znaczenia bez jednego kluczowego elementu: danych, którym można zaufać. Jakość danych jest kamieniem węgielnym cyfryzacji i będziemy do niej powracać w kolejnych częściach.

Nowoczesne firmy budowlane - zwłaszcza te duże - korzystają z dziesiątek, a czasem tysiący różnych systemów i baz danych (rys. 4.2-1). Systemy te muszą być nie tylko regularnie wypełniane nowymi informacjami, ale także efektywnie współpracować ze sobą. Wszystkie nowe dane generowane w wyniku przetwarzania napływających informacji są integrowane w tych środowiskach i służą do rozwiązywania określonych zadań biznesowych.

Podczas gdy w przeszłości decyzje dotyczące konkretnych zadań biznesowych były podejmowane przez menedżerów najwyższego szczebla - tzw. HiPPO (rys. 2.1-9) - na podstawie doświadczenia i intuicji, obecnie, wraz z gwałtownym wzrostem ilości informacji, podejście to staje się kontrowersyjne. Zastępuje je zautomatyzowana analityka, która pracuje z danymi w czasie rzeczywistym.

"Tradycyjno-manualne" dyskusje na poziomie wykonawczym dotyczące procesów biznesowych przesuną się w kierunku analityki operacyjnej, która wymaga szybkich odpowiedzi na zapytania biznesowe.

Era, w której księgowi, brygadziści i kosztorysanci ręcznie generowali raporty i tabele podsumowujące oraz prezentacje danych projektowych przez wiele dni i tygodni, należy już do przeszłości. Obecnie szybkość i terminowość podejmowania decyzji staje się kluczowym czynnikiem przewagi konkurencyjnej.



Rys. 4.2-1 Branża budowlana potrzebuje dni na obliczenia i podjęcie decyzji, w przeciwieństwie do innych branż, w których odbywa się to w ciągu godzin lub minut.

Główna różnicą między branżą budowlaną a bardziej zaawansowanymi cyfrowo branżami (Rys. 4.2-1) jest niski poziom jakości i standaryzacji danych. Przestarzałe podejście do generowania, przesyłania i przetwarzania informacji spowalnia procesy i powoduje chaos. Brak jednolitych standardów jakości danych utrudnia wdrożenie kompleksowej automatyzacji

Jednym z głównych wyzwań pozostaje niska jakość danych wejściowych, a także brak sformalizowanych procesów ich przygotowywania i validacji. Bez wiarygodnych i spójnych danych nie może być mowy o skutecznej integracji między systemami. Prowadzi to do opóźnień, błędów i zwiększonych kosztów na każdym etapie cyklu życia projektu.

W kolejnych rozdziałach książki przyjrzymy się szczegółowo, w jaki sposób można poprawić jakość danych, standaryzować procesy i skrócić drogę od informacji do wysokiej jakości, zweryfikowanych i spójnych danych.

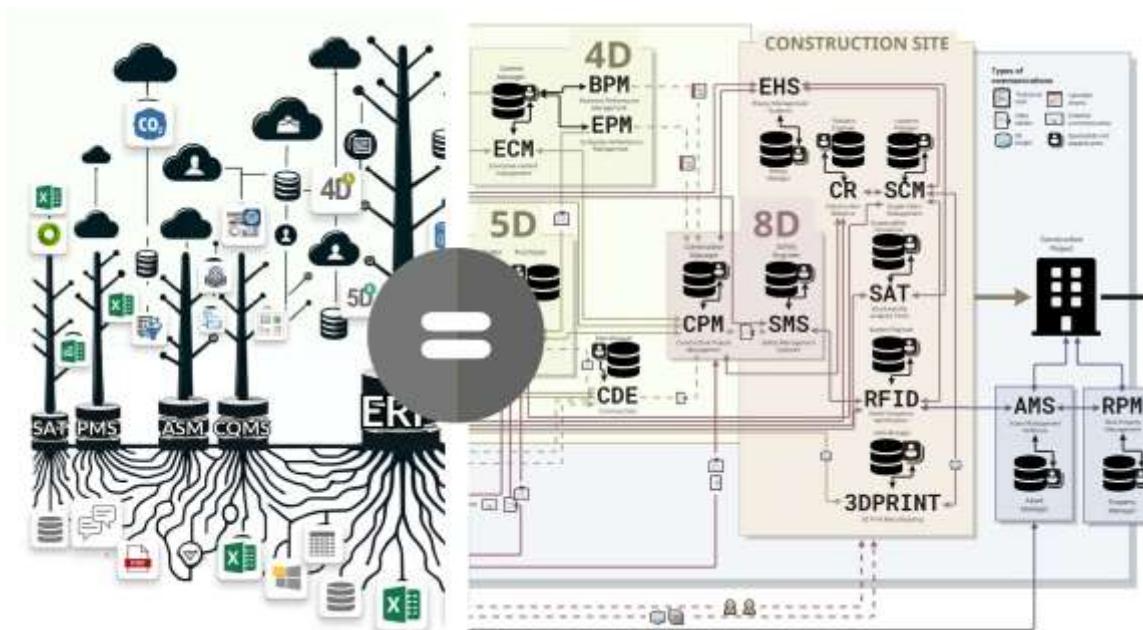
Standaryzacja i integracja

Skuteczne zarządzanie danymi wymaga jasnej strategii standaryzacji. Tylko przy jasnych wymaganiach dotyczących struktury i jakości danych można zautomatyzować validację danych, ograniczyć operacje ręczne i przyspieszyć podejmowanie świadomych decyzji na wszystkich etapach projektu.

W codziennej praktyce firma budowlana musi codziennie przetwarzać setki plików: wiadomości e-mail, dokumenty PDF, pliki projektowe CAD, dane z czujników IoT, które muszą być zintegrowane z procesami biznesowymi firmy.

Las ekosystemu baz danych i narzędzi firmy (Rys. 4.2-2) musi nauczyć się czerpać składniki odżywcze z napływających danych wieloformatowych, aby osiągnąć pożądane przez firmę wyniki.

Aby skutecznie radzić sobie z przepływem danych, niekoniecznie trzeba zatrudniać armię menedżerów, należy najpierw opracować ścisłe wymagania i standardy dotyczące danych oraz użyć odpowiednich narzędzi do ich automatycznej walidacji, harmonizacji i przetwarzania.



Rys. 4.2-2 Zapewnienie zdrowego ekosystemu firmy wymaga wysokiej jakości i terminowego zasilania jej systemów.

Aby zautomatyzować proces walidacji i harmonizacji danych (w celu późniejszej automatycznej integracji), należy rozpocząć od opisania minimalnych niezbędnych wymagań dotyczących danych dla każdego konkretnego systemu. Wymagania te definiują:

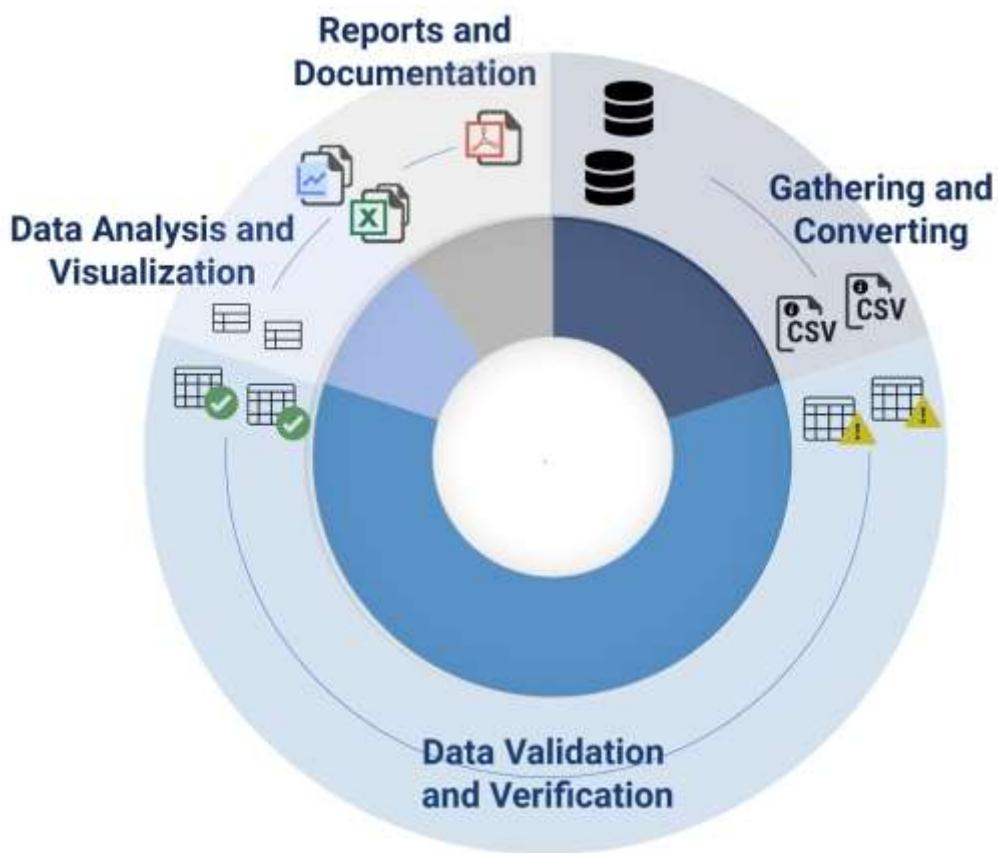
- Co dokładnie musisz zdobyć?
- W jakiej formie (struktura, format)?
- Jakie atrybuty są obowiązkowe?
- Jakie tolerancje w zakresie dokładności i kompletności są dopuszczalne?

Wymagania dotyczące danych opisują kryteria jakości, struktury i kompletności otrzymywanych i przetwarzanych informacji. Na przykład w przypadku tekstów w dokumentach PDF ważne jest zapewnienie dokładnego formatowania zgodnie ze standardami branżowymi (rys. 7.2-14 - rys. 7.2-16). Obiekty w modelach CAD muszą mieć prawidłowe atrybuty (wymiary, kody, linki do klasyfikatorów) (rys. 7.3-9, rys. 7.3-10). W przypadku skanowania umów ważne są wyraźne daty i możliwość automatycznego wyodrębnienia kwoty i kluczowych warunków (rys. 4.1-7 - rys. 4.1-10).

Formułowanie wymagań dotyczących danych i automatyczne sprawdzanie ich zgodności jest jednym z najbardziej czasochłonnych, ale krytycznych kroków. Jest to najbardziej czasochłonny krok w procesach biznesowych.

Jak wspomniano w części 3 tej książki, od 50% do 90% czasu specjalistów ds. analityki biznesowej (BI) poświęca się na przygotowanie danych, a nie na ich analizę (Rys. 3.2-5). Proces ten obejmuje gromadzenie, weryfikację, walidację, harmonizację i strukturyzację danych.

Według ankiety przeprowadzonej w 2016 roku [95], naukowcy zajmujący się danymi w wielu różnych dziedzinach o szerokim spektrum stwierdzili, że większość czasu pracy (około 80%) spędzają na robieniu tego, co najmniej lubią robić (Rys. 4.2-3): gromadzeniu istniejących zbiorów danych i ich organizowaniu (ujednolicaniu, strukturyzowaniu). W związku z tym mniej niż 20% czasu pozostaje na kreatywne zadania, takie jak znajdowanie wzorców i prawidłowości, które doprowadzą do nowych spostrzeżeń i odkryć.



Rys. 4.2-3 Weryfikacja i zapewnienie jakości danych jest najbardziej kosztownym, czasochłonnym i złożonym etapem przygotowania danych do integracji z innymi systemami.

Skuteczne zarządzanie danymi w firmie budowlanej wymaga kompleksowego podejścia, które obejmuje parametryzację zadań, formułowanie wymagań dotyczących jakości danych i stosowanie odpowiednich narzędzi do ich automatycznej walidacji.

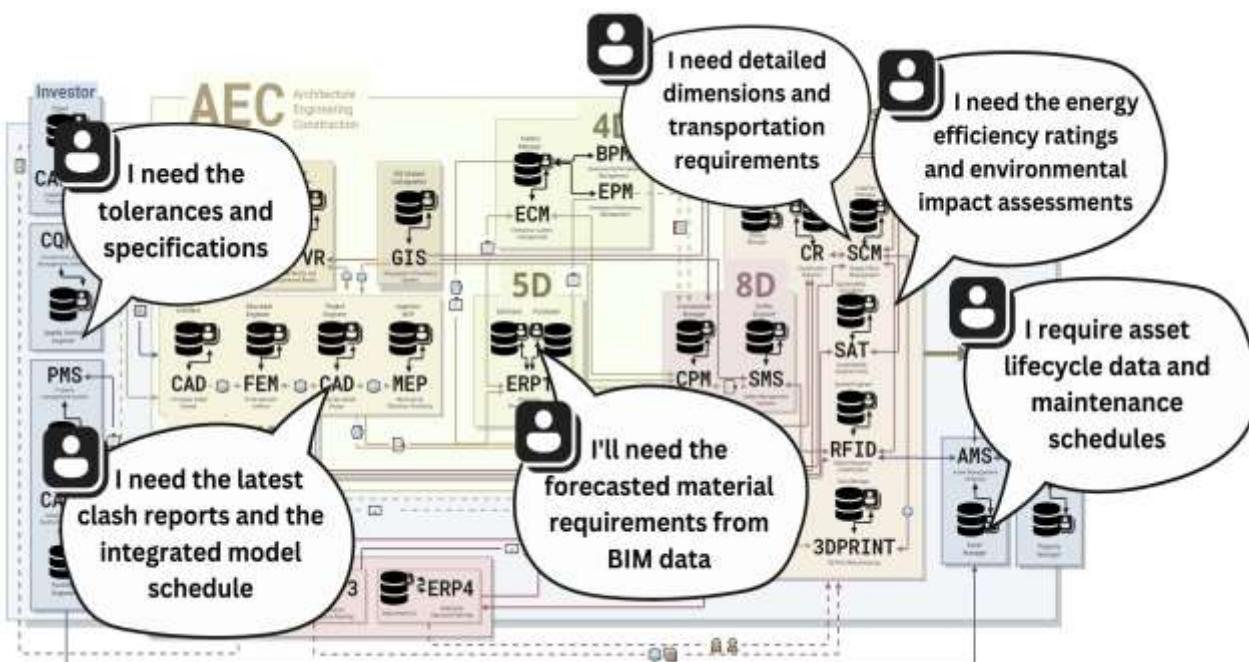
Interoperacyjność cyfrowa zaczyna się od wymagań

Wraz ze wzrostem liczby systemów cyfrowych w firmach rośnie potrzeba spójności danych między nimi. Menedżerowie odpowiedzialni za różne systemy IT często nie są w stanie nadążyć za rosnącą ilością informacji i różnorodnością formatów. W takich okolicznościach są zmuszeni prosić specjalistów o tworzenie danych w formie odpowiedniej do wykorzystania w innych aplikacjach i platformach.

To z kolei wymaga od inżynierów i personelu generującego dane dostosowania się do wielu wymagań, często bez przejrzystości i jasnego zrozumienia, gdzie i w jaki sposób dane będą wykorzystywane w przyszłości. Brak znormalizowanego podejścia do przetwarzania informacji prowadzi do nieefektywności i zwiększych kosztów na etapie weryfikacji, która często jest wykonywana ręcznie ze względu na złożoność i niestandardowy charakter danych.

Kwestia standaryzacji danych to nie tylko kwestia wygody czy automatyzacji. Jest to bezpośrednia strata finansowa. Według raportu IBM z 2016 r. roczna strata wynikająca z niskiej jakości danych w USA wynosi 3,1 biliona dolarów [96]. Ponadto badania przeprowadzone przez MIT i inne firmy konsultingowe pokazują, że koszt niskiej jakości danych może wynosić nawet 15-25% przychodów firmy [97].

W takich warunkach krytyczne staje się posiadanie jasno zdefiniowanych wymagań dotyczących danych i opisów tego, jakie parametry, w jakim formacie i na jakim poziomie szczegółowości powinny być zawarte w tworzonych obiektach. Bez sformalizowania tych wymagań niemożliwe jest zagwarantowanie jakości danych i kompatybilności między systemami i fazami projektu (rys. 4.2-4).

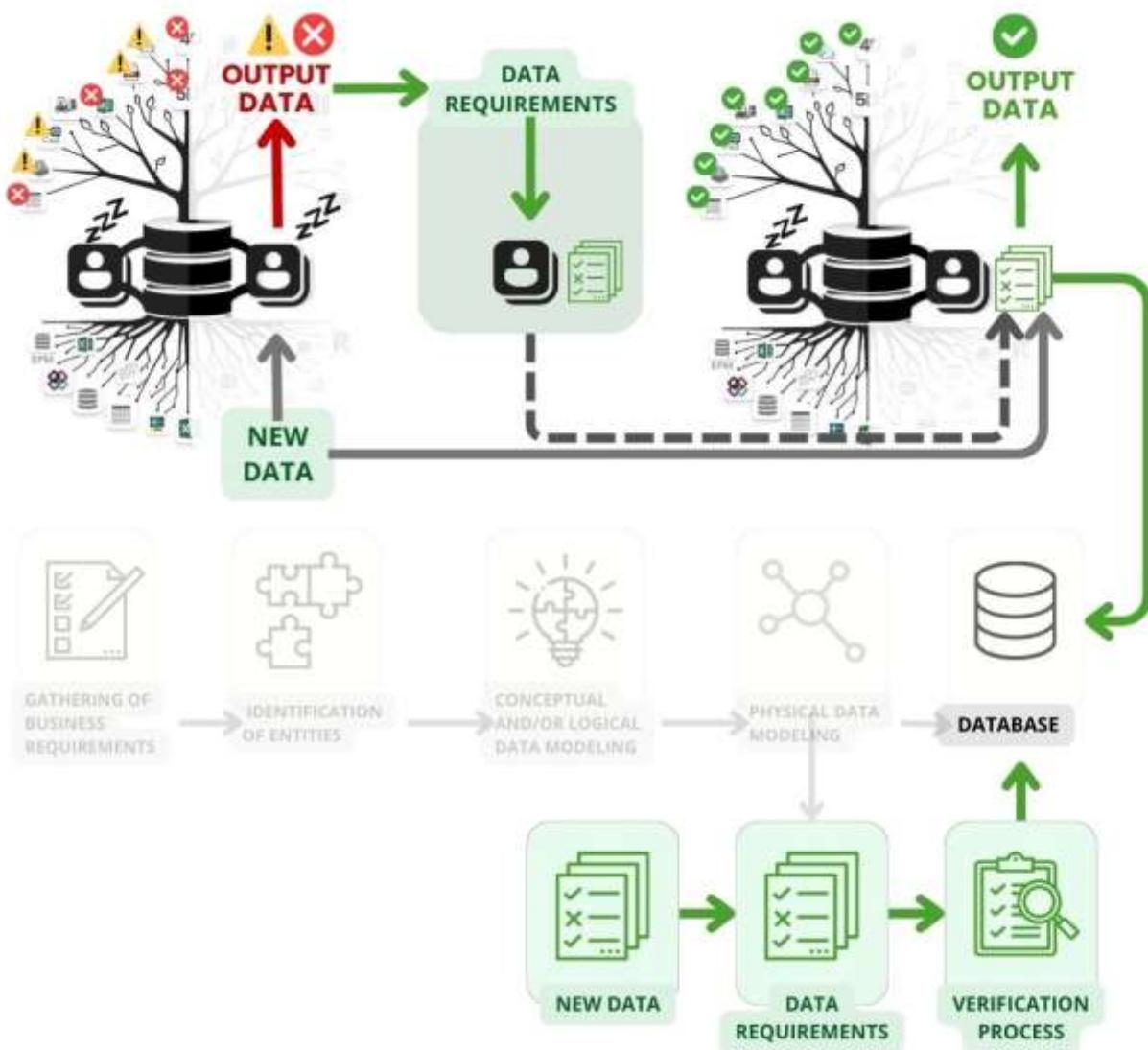


Rys. 4.2-4 Firma opiera się na interakcji różnych ról, z których każda wymaga pewnych parametrów i wartości, które są krytyczne dla realizacji celów biznesowych.

Aby sformułować prawidłowe wymagania dotyczące danych, należy zrozumieć procesy biznesowe na poziomie danych. Projekty budowlane różnią się rodzajem, skalą i liczbą uczestników, a każdy system - czy

to modelowanie (CAD (BIM)), harmonogramowanie (ERP 4D), kalkulacja kosztów (ERP 5D) czy logistyka (SCM) - wymaga własnych unikalnych parametrów dla danych wejściowych (wejściowych encji-elementów).

W zależności od tych potrzeb menedżerowie biznesowi muszą albo zaprojektować nowe struktury danych, aby spełnić wymagania, albo dostosować istniejące tabele i bazy danych. Jakość utworzonych danych będzie bezpośrednio zależeć od tego, jak precyzyjnie i poprawnie sformułowane są wymagania (rys. 4.2-5).



Rys. 4.2-5 Jakość danych zależy od jakości wymagań, które są tworzone dla konkretnych przypadków użycia danych.

Ponieważ każdy system ma swoje specyficzne wymagania dotyczące danych, pierwszym krokiem w formułowaniu ogólnych wymagań jest kategoryzacja wszystkich elementów zaangażowanych w procesy biznesowe. Oznacza to konieczność podziału obiektów na klasy i grupy klas odpowiadające konkretnym

systemom lub zadaniom aplikacji. Dla każdej takiej grupy opracowywane są oddzielne wymagania dotyczące struktury, atrybutów i jakości danych.

W praktyce jednak wdrożenie tego podejścia napotyka na poważne wyzwanie: brak wspólnego języka grupowania danych. Rozbieżne klasyfikacje, zduplikowane identyfikatory i niekompatybilne formaty powodują, że każda firma, każde oprogramowanie, a nawet każdy projekt tworzą własne, odizolowane modele i klasy danych. W rezultacie powstaje cyfrowa "wieża Babel", w której przenoszenie informacji między systemami wymaga wielokrotnej konwersji do odpowiednich modeli i klas danych, często wykonywanej ręcznie. Barierę tę można pokonać jedynie poprzez przejście na uniwersalne klasyfikatory i ustandaryzowane zestawy wymagań.

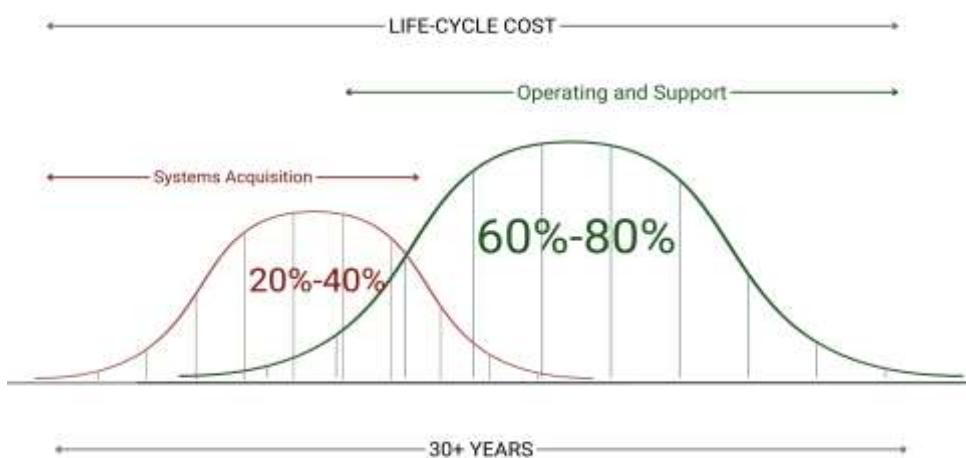
Wspólny język konstrukcji: rola klasyfikatorów w transformacji cyfrowej

W kontekście cyfryzacji i automatyzacji procesów kontroli i przetwarzania, systemy klasyfikacji odgrywają szczególną rolę - rodzaj "cyfrowych słowników", które zapewniają jednolitość w opisie i parametryzacji obiektów. Klasyfikatory tworzą "wspólny język", który umożliwia grupowanie danych według ich znaczenia i integrację między różnymi systemami, poziomami zarządzania i fazami cyklu życia projektu.

Najbardziej namacalny wpływ klasyfikatorów dotyczy ekonomii cyklu życia budynku, gdzie najważniejszym aspektem jest optymalizacja długoterminowych kosztów operacyjnych. Badania pokazują, że koszty operacyjne stanowią do 80% całkowitego kosztu posiadania budynku, który jest trzykrotnie wyższy niż początkowe koszty budowy (rys. 4.2-6) [98]. Oznacza to, że decyzja o przyszłych kosztach jest w dużej mierze podejmowana na etapie projektowania

Dlatego wymagania od inżynierów operacyjnych (CAFM, AMS, PMS, RPM) powinny stać się punktem wyjścia do generowania wymagań dotyczących danych w fazie projektowania (rys. 1.2-4). Systemy te nie powinny być postrzegane jako końcowy etap projektu, ale jako integralna część całego cyfrowego ekosystemu projektu, od koncepcji do demontażu

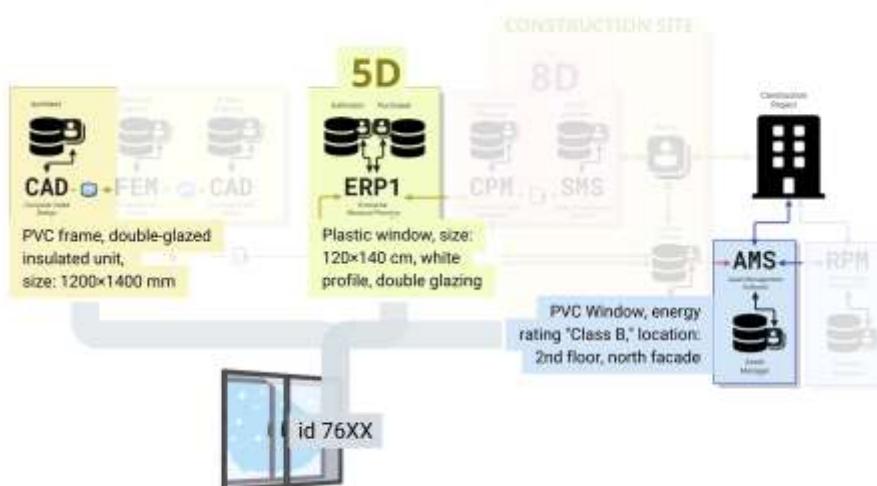
Nowoczesny klasyfikator to nie tylko system kodów do grupowania. Jest to mechanizm wzajemnego zrozumienia między architektami, inżynierami, kosztorysantami, logistykami, konserwatorami i systemami informatycznymi. Tak jak autopilot samochodu musi jednoznacznie i dokładnie rozpoznawać obiekty drogowe, tak cyfrowe systemy budowlane i ich użytkownicy muszą jednoznacznie interpretować ten sam element projektu dla różnych systemów poprzez klasę elementu.



Rys. 4.2-6 Koszty eksploatacji i utrzymania trzykrotnie przekraczają koszty budowy, stanowiąc 60-80% całkowitych kosztów cyklu życia budynku (zgodnie z [99]).

Poziom rozwoju klasyfikatorów bezpośrednio koreluje z głębokością cyfryzacji firmy i jej dojrzałością cyfrową. Organizacje o niskim poziomie dojrzałości cyfrowej borykają się z rozdrobnionymi danymi, niekompatybilnymi systemami informatycznymi, a w rezultacie z niekompatybilnymi i nieefektywnymi klasyfikatorami. W takich firmach ten sam element może często mieć różne identyfikatory grupowania w różnych systemach, co znacznie utrudnia ostateczną integrację i uniemożliwia automatyzację procesów.

Na przykład to samo okno w projekcie może być różnie oznaczone w modelu CAD, systemie szacowania i konserwacji (rys. 4.2-7) ze względu na wielowymiarowe postrzeganie elementów przez różnych uczestników procesu. Dla kosztorysanta w elemencie kategorii okna ważna jest objętość i koszt, dla służb utrzymania - dostępność i łatwość konserwacji, dla architekta - cechy estetyczne i funkcjonalne. W rezultacie ten sam element może wymagać różnych parametrów.



Rys. 4.2-7 Przy niespójnej klasyfikacji między systemami, element straci pewne informacje o atrybutach na każdym etapie przejścia do innego systemu.

Ze względu na trudności w jednoznaczny zdefiniowaniu klasyfikacji elementów budynków, specjalisci z różnych dziedzin często przypisują niezgodne klasy do tego samego elementu. Prowadzi to do utraty

ujednoliconego widoku obiektu, co wymaga późniejszej ręcznej interwencji w celu zharmonizowania różnych systemów klasyfikacji i ustanowienia spójności między typami i klasami zdefiniowanymi przez różnych specjalistów.

W wyniku tej niespójności dokumentacja operacyjna otrzymywana przez dział zaopatrzenia (ERP), gdy element budowlany jest kupowany od producenta, często nie może być prawidłowo powiązana z klasyfikacją tego elementu na placu budowy (PMIS, SCM). W konsekwencji krytyczne informacje prawdopodobnie nie zostaną zintegrowane z systemami zarządzania infrastrukturą i aktywami (CAFM, AMS), co stwarza poważne problemy podczas uruchamiania, a także podczas późniejszej konserwacji (AMS, RPM) lub wymiany elementu.

W firmach o wysokiej dojrzałości cyfrowej klasyfikatory odgrywają rolę systemu nerwowego, który łączy wszystkie przepływy informacji. Ten sam element otrzymuje unikalny identyfikator, dzięki czemu może być przenoszony między systemami CAD, ERP, AMS i CAFM i ich klasyfikatorami bez zniekształceń lub strat.

Aby zbudować skuteczne klasyfikatory, należy zrozumieć, w jaki sposób wykorzystywane są dane. Ten sam inżynier może nazwać i sklasyfikować element inaczej w różnych projektach. Tylko poprzez gromadzenie statystyk użytkowania na przestrzeni lat można opracować stabilny system klasyfikacji. Pomaga w tym uczenie maszynowe: algorytmy analizują tysiące projektów (rys. 9.1-10), identyfikując prawdopodobne klasy i parametry za pomocą uczenia maszynowego (rys. 10.1-6). Automatyczna klasyfikacja jest szczególnie cenna w środowiskach, w których ręczna klasyfikacja nie jest możliwa ze względu na ilość danych. Systemy automatycznej klasyfikacji będą w stanie rozróżnić podstawowe kategorie na podstawie minimalnie wypełnionych parametrów pozycji (więcej szczegółów w dziewiątej i dziesiątej części książki).

Opracowane systemy klasyfikatorów stają się katalizatorami dalszej cyfryzacji, tworząc podstawę do jej rozwoju:

- Zautomatyzowane szacowanie kosztów i harmonogramów projektów.
- Analiza predykcyjna potencjalnych zagrożeń i konfliktów
- Optymalizacja procesów zaopatrzenia i łańcuchów logistycznych
- Tworzenie cyfrowych bliźniaków budynków i struktur
- Integracja z systemami inteligentnego miasta i Internetu rzeczy

Czas na transformację jest ograniczony - wraz z rozwojem uczenia maszynowego i technologii wizji komputerowej, problem automatycznej klasyfikacji, który był nierozwiązywalny przez dziesięciolecia, zostanie rozwiązany w nadchodzących latach, a firmy budowlane i projektowe, które nie dostosują się na czas, ryzykują powtórzenie losu flot taksówek wypartych przez platformy cyfrowe.

Automatyzacja obliczeń kosztów i harmonogramów, a także duże zbioru danych i uczenie maszynowe zostaną omówione bardziej szczegółowo w piątej i dziewiątej części książki. Ryzyko powtórzenia się losu flot taksówkowych i uberyzacji branży budowlanej omówiono szczegółowo w dziesiątej części książki.

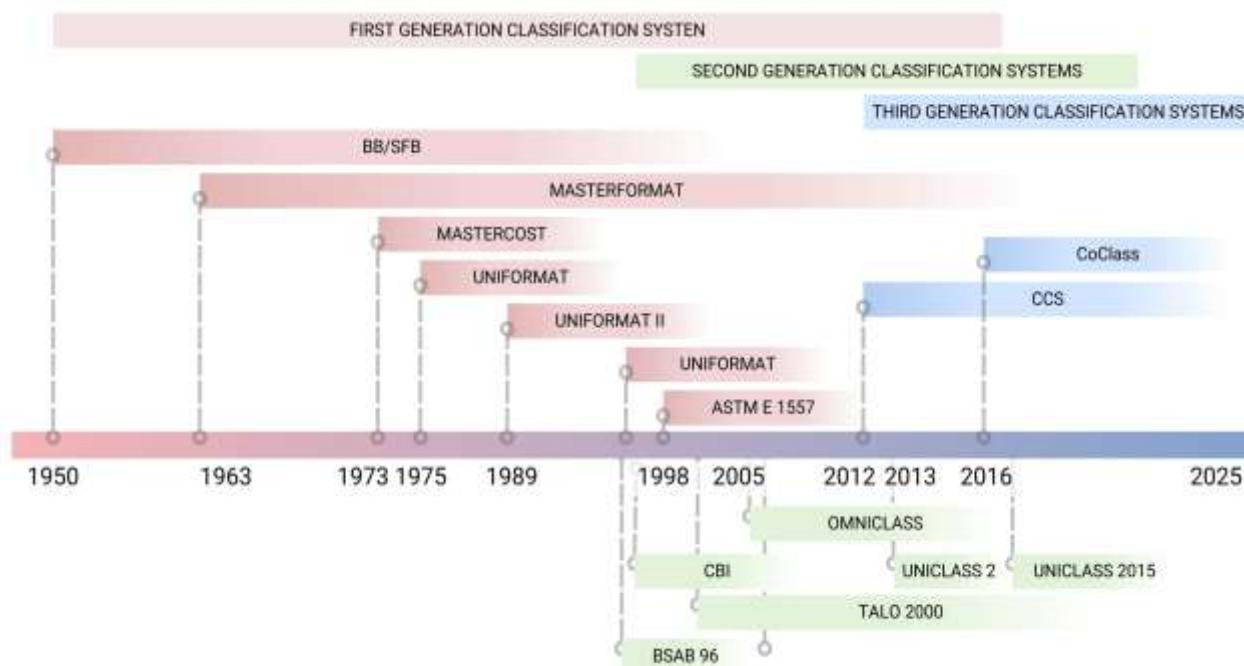
Zrozumienie kluczowej roli klasyfikatorów w cyfrowej transformacji branży budowlanej wymaga sięgnięcia

do historii ich ewolucji. To właśnie kontekst historyczny pozwala nam uświadomić sobie, jak ewoluowały podejścia do klasyfikacji i jakie trendy definiują ich obecny stan.

Masterformat, OmniClass, Uniclass i CoClass: ewolucja systemów klasyfikacji

Historycznie rzecz biorąc, klasyfikatory elementów konstrukcyjnych i pracy ewoluowały w trzech generacjach, z których każda odzwierciedlała poziom dostępnej technologii i bieżące potrzeby branży w danym okresie (Rys. 4.2-8):

- **Pierwsza generacja** (wczesne lata 50. do późnych lat 80.) - katalogi papierowe, hierarchiczne klasyfikatory używane lokalnie (np. Masterformat, SfB).
- **Druga generacja** (od końca lat 90. do połowy 2010 r.) to arkusze kalkulacyjne i ustrukturyzowane bazy danych zaimplementowane w programach Excel i Access (ASTM E 1557, OmniClass, Uniclass 1997).
- **Trzecia generacja** (lata 2010 do chwili obecnej) - usługi cyfrowe i interfejsy API -interfejsy, integracja z CAD (BIM), automatyzacja (Uniclass 2015, CoClass).



Rys. 4.2-8 Trzy generacje klasyfikatorów branży budowlanej.

W ciągu ostatnich dziesięcioleci nastąpiło zmniejszenie złożoności hierarchicznej (rys. 4.2-9) klasyfikatorów: podczas gdy wczesne systemy, takie jak OmniClass, wykorzystywały do 7 poziomów zagłędzienia do opisania 6887 klas, nowoczesne rozwiązania, takie jak CoClass, są ograniczone do 3 poziomów z 750 klasami. Ułatwia to pracę z danymi przy jednoczesnym zachowaniu niezbędnej szczegółowości. Uniclass 2015, często używany jako standard w Wielkiej Brytanii, łączy 7210 klas na zaledwie 4 poziomach, dzięki czemu jest wygodny dla projektów CAD i zamówień publicznych.

Classifier	Table / Objects	Number of classes	Nesting depth
OmniClass	Table 23 Products	6887	7 levels
Uniclass 2015	Pr — Products	7210	4 levels
CoClass, CCS	Components	750	3 levels

Rys. 4.2-9 Z każdą nową generacją klasyfikatorów złożoność kategoryzacji spada skokowo.

W systemach szacowania konstrukcji w różnych krajach, ze względu na różnice w klasifikacjach, nawet typowy element, taki jak betonowa ściana fundamentowa, może być opisany na bardzo różne sposoby (Rys. 4.2-10). Różnice te odzwierciedlają krajowe praktyki budowlane, stosowane systemy pomiarowe, podejścia do klasyfikacji materiałów oraz wymogi prawne i techniczne obowiązujące w każdym kraju.



Rys. 4.2-10 Ten sam element jest używany w projektach w różnych krajach poprzez różne opisy i klasyfikacje.

Różnorodność klasyfikacji tych samych elementów komplikuje współpracę międzynarodową i sprawia, że porównanie kosztów i zakresu prac w ramach międzynarodowych projektów jest czasochłonne, a czasem niemal niemożliwe. Obecnie nie ma jednego uniwersalnego klasyfikatora na poziomie globalnym - każdy kraj lub region opracowuje własne systemy oparte na lokalnych normach, języku i kulturze biznesowej:

- **CCS (Dania):** Cost Classification System - system klasyfikacji kosztów w całym cyklu życia obiektu (projektowanie, budowa, eksploatacja). Nacisk kładziony jest na logikę eksploatacji i utrzymania, ale obejmuje również zarządzanie budżetem i zasobami.
- **NS 3451 (Norwegia):** kategoryzuje obiekty według funkcji, elementów projektu i etapów cyklu życia. Używany do zarządzania projektami, szacowania kosztów i planowania długoterminowego.
- **MasterFormat (USA):** system strukturyzacji specyfikacji budowlanych w sekcje (np. beton, elektryka, wykończenie). Koncentruje się na dyscyplinach i rodzajach prac, a nie na elementach funkcjonalnych (w przeciwnieństwie do UniFormat).
- **Uniclass 2 (UK):** jeden z najbardziej szczegółowych klasyfikatorów, wykorzystywany w zamówieniach publicznych i projektach BIM. Ujednolica dane dotyczące obiektów, robót, materiałów i przestrzeni w

jednym systemie.

- **OmniClass:** międzynarodowy standard (opracowany przez CSI w USA) do zarządzania informacjami o obiektach, od bibliotek komponentów po specyfikacje elektroniczne. Odpowiedni do długoterminowego przechowywania danych, kompatybilny z CAD (BIM) i innymi narzędziami cyfrowymi.
- **COBie:** Construction-Operation Building Information Exchange to międzynarodowy standard wymiany danych między fazami projektowania, budowy i eksploatacji. Zawarty w normie BS 1192-4:2014 jako część koncepcji "BIM - model gotowy do użycia". Koncentruje się na transferze informacji (np. specyfikacje sprzętu, gwarancje, kontakty z wykonawcami).

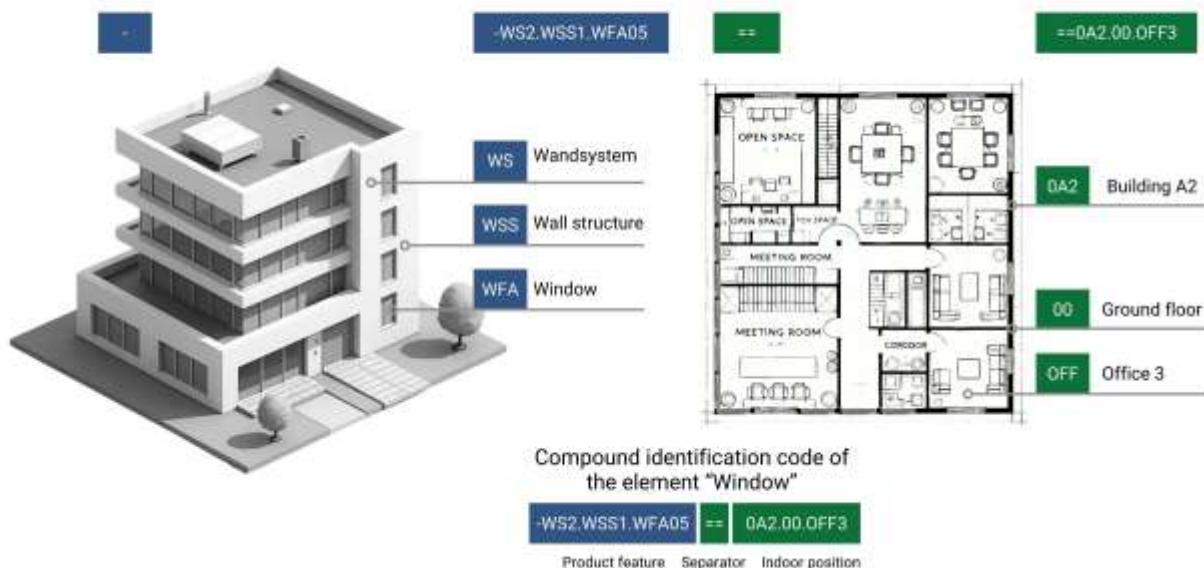
Globalizacja branży budowlanej prawdopodobnie doprowadzi do stopniowej unifikacji systemów klasyfikacji elementów budowlanych, co znacznie zmniejszy zależność od lokalnych norm krajowych. Proces ten może rozwijać się zgodnie z ewolucją komunikacji internetowej, gdzie uniwersalne protokoły przesyłania danych ostatecznie wyparły rozbieżne formaty lokalne, zapewniając globalną interoperacyjność systemów.

Alternatywną ścieżką rozwoju może być bezpośrednie przejście do systemów automatycznej klasyfikacji opartych na technologiach uczenia maszynowego. Technologie te, które są obecnie rozwijane głównie w dziedzinie transportu autonomicznego, mają znaczny potencjał do zastosowania w dużych zbiorach danych projektowych CAD (rys. 10.1-6).

Obecnie sytuacja nie ogranicza się jedynie do krajowych klastrów klasyfikatorów. Ze względu na wiele osobliwości nieuwzględnionych na poziomie krajowym, każda firma musi ujednolicić i znormalizować kategorie elementów i zasobów, z którymi pracuje.

Z reguły proces ten rozpoczyna się na małą skalę - od lokalnych tabel obiektów lub wewnętrznych systemów etykietowania. Jednak celem strategicznym jest przejście do wspólnego języka opisu wszystkich elementów, który byłby zrozumiały nie tylko w firmie, ale także poza nią - najlepiej zharmonizowany z międzynarodowymi lub branżowymi klasyfikatorami (rys. 4.2-8). Takie podejście ułatwia integrację z partnerami zewnętrznymi, systemami cyfrowymi i promuje tworzenie ujednoliconych procesów end-to-end w ramach cyklu życia obiektów.

Przed przejściem na automatyzację i skalowalne systemy informatyczne konieczne jest albo użycie klasyfikatorów na poziomie krajowym, albo zbudowanie własnej, logicznej i jednoznacznej struktury identyfikacji elementów. Każdy obiekt - czy to okno (Rys. 4.2-11), drzwi czy system inżynierijny - musi być opisany w taki sposób, aby można go było jednoznacznie rozpoznać w dowolnym systemie cyfrowym firmy. Ma to kluczowe znaczenie przy przejściu od płaskich rysunków do modeli cyfrowych, obejmujących zarówno fazę projektowania, jak i eksploatację budynku.



Rys. 4.2-11 Przykład identyfikatora złożonego elementu budynku okna na podstawie klasyfikacji i pozycji w budynku.

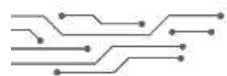
Jednym z przykładów wewnętrznych klasyfikatorów może być opracowanie złożonego kodu identyfikacyjnego (Rys. 4.2-11). Taki kod łączy w sobie kilka poziomów informacji: cel funkcjonalny elementu (np. "okno w ścianie"), jego typ i dokładne odniesienie przestrzenne - budynek A2, piętro 0, pomieszczenie 3. Taka wielopoziomowa struktura pozwala na stworzenie jednolitego systemu nawigacji po cyfrowych modelach i dokumentacji, szczególnie na etapach weryfikacji i transformacji danych, gdzie wymagane jest jednoznaczne grupowanie elementów. Jednoznaczne rozpoznawanie elementów zapewnia spójność między działami i zmniejsza ryzyko duplikacji, błędów i utraty informacji.

Dobrze zbudowany klasyfikator to nie tylko dokument techniczny, to podstawa cyfrowego ekosystemu firmy:

- zapewnia zgodność danych między systemami;
- zmniejsza koszty wyszukiwania i przetwarzania informacji;
- zwiększa przejrzystość i łatwość zarządzania;
- tworzy podstawę do skalowania i automatyzacji.

Znormalizowany opis obiektów, poprzez wykorzystanie krajowych klasyfikatorów lub zastrzeżonych złożonych kodów identyfikacyjnych, staje się podstawą spójnych danych, niezawodnej wymiany informacji i późniejszego wdrażania inteligentnych usług - od zautomatyzowanych zamówień po cyfrowe bliźniaki.

Po zakończeniu etapu strukturyzacji danych wieloformatowych i wybraniu klasyfikatora, który będzie używany do rozpoznawania i grupowania elementów, kolejnym krokiem jest prawidłowe modelowanie danych. Proces ten obejmuje identyfikację kluczowych parametrów, zbudowanie logicznej struktury danych



i opisanie relacji między elementami.

ROZDZIAŁ 3.3.

MODELOWANIE DANYCH I CENTRUM DOSKONAŁOŚCI

Modelowanie danych: model koncepcyjny, logiczny i fizyczny

Efektywne zarządzanie danymi (ustrukturyzowanymi i skategoryzowanymi przez nas wcześniej) nie jest możliwe bez przemyślanej struktury ich przechowywania i przetwarzania. Aby zapewnić dostęp i spójność informacji na etapach przechowywania i przetwarzania, firmy wykorzystują modelowanie danych, metodologię, która pozwala im projektować tabele, bazy danych i powiązania między nimi zgodnie z wymaganiami biznesowymi.

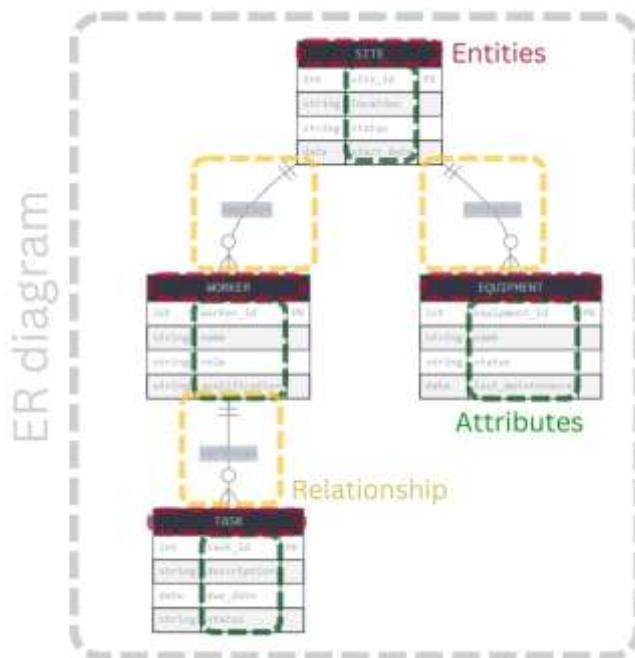
Modelowanie danych jest fundamentem, na którym budowany jest każdy cyfrowy ekosystem. Bez opisu systemów, wymagań i modelowania danych, inżynierowie i specjalisci tworzący dane nie wiedzą lub nie rozumieją, gdzie tworzone przez nich dane będą wykorzystywane.

Podobnie jak w przypadku budowy budynku, gdzie nie można rozpocząć układania cegieł bez planu, tworzenie systemu hurtowni danych wymaga jasnego zrozumienia, jakie dane będą wykorzystywane, w jaki sposób będą one połączone i kto będzie z nimi pracował. Bez opisu procesów i wymagań inżynierowie i specjalisci tworzący dane tracą z oczu to, gdzie i w jaki sposób dane będą wykorzystywane w przyszłości.

Model danych służy jako pomost między biznesem a IT. Umożliwia formalizację wymagań, strukturyzację informacji i ułatwia komunikację między interesariuszami. W tym sensie modelowanie danych jest podobne do pracy architekta, który zgodnie z planem klienta opracowuje plan budynku, a następnie przekazuje go budowniczym - administratorom baz danych i programistom - w celu wdrożenia (utworzenia bazy danych).

Tak więc każda firma budowlana, oprócz strukturyzowania i kategoryzowania elementów i zasobów (rys. 4.2-11), musi opanować sztukę "budowania" baz danych (tabel) i nauczyć się tworzyć powiązania między nimi, jakby łącząc cegły w niezawodny i mocny mur wiedzy z danych firmy. Kluczowe pojęcia w modelowaniu danych (Rys. 4.3-1) obejmują:

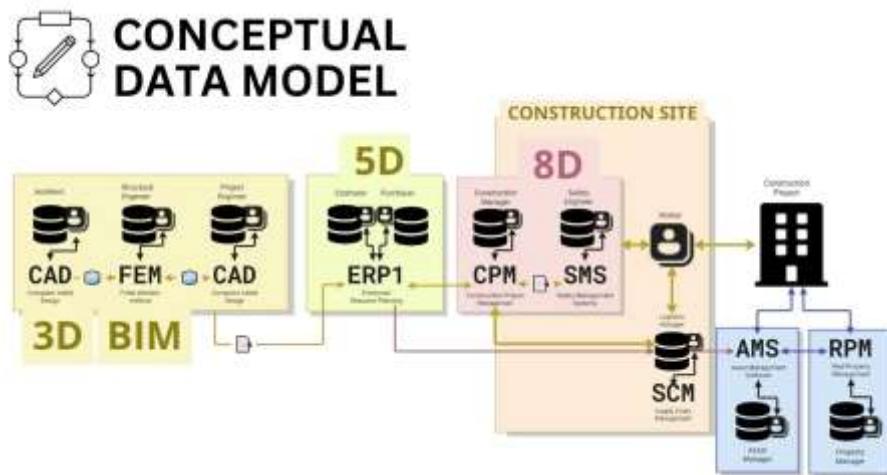
- **Podmioty** to obiekty, o których należy gromadzić dane. We wczesnej fazie projektowania encja może być pojedynczym elementem (np. "drzwi"), a w modelu szacunkowym może być grupą skategoryzowanych elementów (np. "drzwi wewnętrzne").
- **Atrybuty** to cechy podmiotów, które opisują ważne szczegóły: wymiary, właściwości, koszty montażu, logistykę i inne parametry.
- **Relacje (powiązania)** - pokazują, w jaki sposób jednostki oddziałują na siebie nawzajem. Mogą być jednego z typów: "jeden do jednego", "wiele do jednego", "wiele do wielu".
- Diagramy **ER** (Entity-Relationship diagrams) to diagramy wizualne, które pokazują encje, atrybuty i relacje między nimi. Diagramy ER mogą być koncepcyjne, logiczne i fizyczne - każdy z nich odzwierciedla inny poziom szczegółowości.



Rys. 4.3-1 Diagram ER konceptualnej struktury bazy danych z encjami, atrybutami i relacjami.

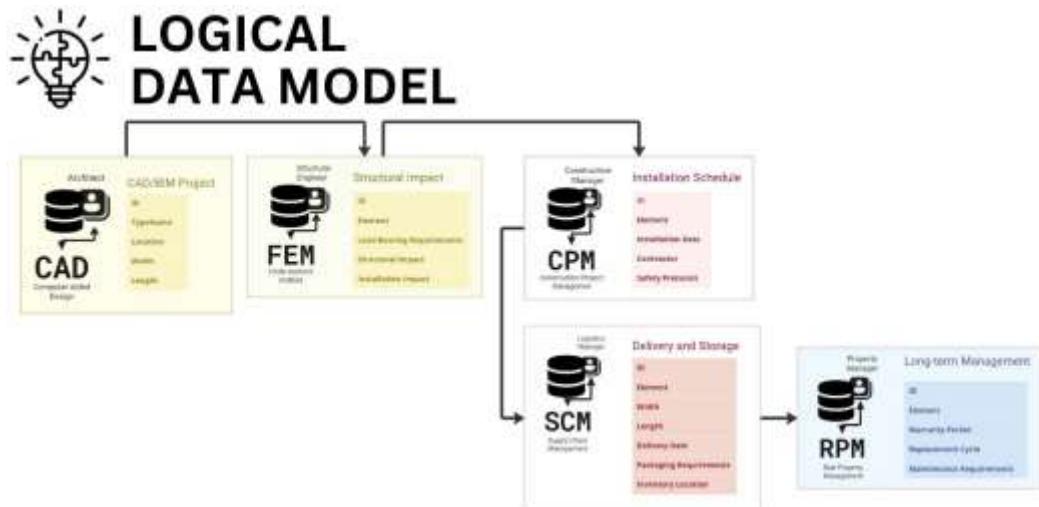
Proces projektowania danych i definiowania relacji między nimi tradycyjnie dzieli się na trzy główne modele. Każdy z nich spełnia określone funkcje, różniąc się poziomem szczegółowości i stopniem abstrakcji w reprezentowaniu struktury danych:

- **Koncepcyjny model danych:** model ten opisuje główne jednostki i ich relacje bez wchodzenia w szczegóły atrybutów. Jest on zwykle używany na początkowych etapach planowania. Na tym etapie możemy szkicować z baz danych i systemów, aby pokazać relacje między różnymi działami i specjalistami.



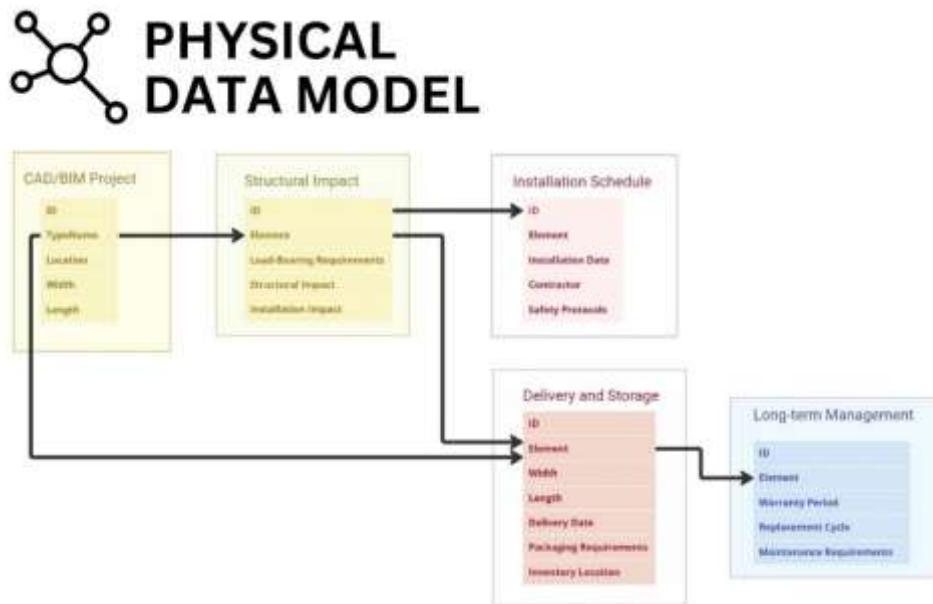
Rys. 4.3-2 Schemat koncepcyjny opisuje zawartość systemu: wysokopoziomowa reprezentacja relacji, bez szczegółów technicznych.

- **Logiczny model danych:** W oparciu o model koncepcyjny, logiczny model danych zawiera szczegółowe opisy jednostek, atrybutów, kluczy i relacji, mapując informacje biznesowe i reguły.



Rys. 4.3-3 Logiczny model danych zawiera szczegółowe informacje na temat typów danych, relacji i kluczy, ale bez implementacji systemu.

- **Fizyczny model danych:** Model ten opisuje struktury niezbędne do implementacji bazy danych, w tym tabele, kolumny i relacje. Koncentruje się na wydajności bazy danych, strategiach indeksowania i fizycznej pamięci masowej w celu optymalizacji fizycznego wdrażania baz danych.

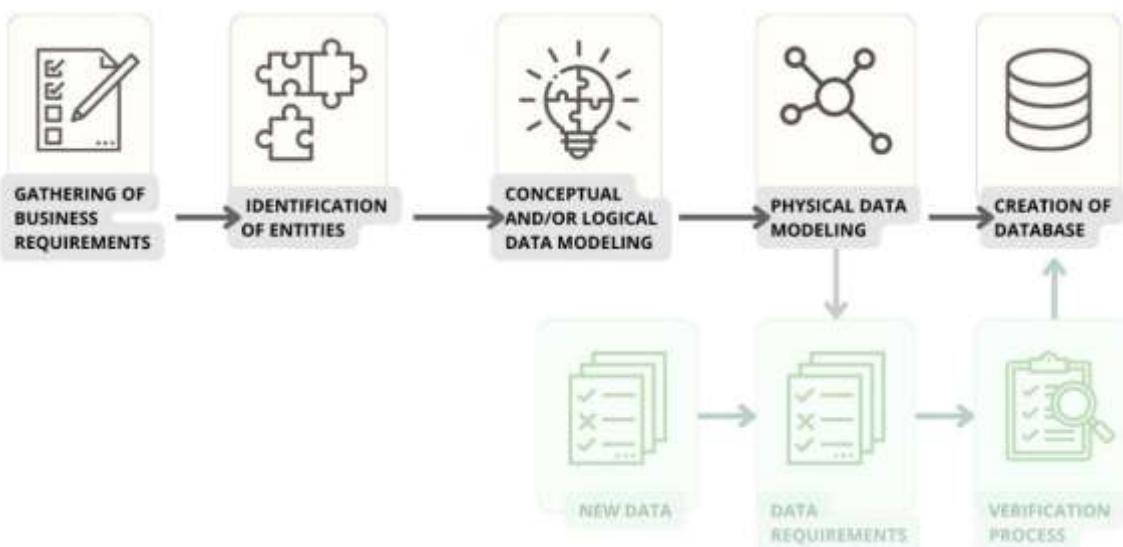


Rys. 4.3-4 Fizyczny model danych definiuje sposób implementacji systemu, w tym tabele i konkretne szczegóły bazy danych.

Podczas projektowania baz danych i projektowania relacji tabelarycznych, zrozumienie poziomów abstrakcji odgrywa kluczową rolę w budowaniu efektywnej architektury systemu.

Skuteczna metodologia modelowania danych pozwala łączyć cele biznesowe z implementacją techniczną, dzięki czemu cały łańcuch procesów jest bardziej przejrzysty i łatwiejszy w zarządzaniu. Modelowanie danych nie jest jednorazowym zadaniem, ale procesem obejmującym kolejne kroki (Rys. 4.3-5):

- **Zbieranie wymagań biznesowych:** definiowane są kluczowe zadania, cele i przepływy informacji. Jest to etap aktywnej interakcji z ekspertami i użytkownikami.
- **Identyfikacja podmiotów:** wyróżniono główne obiekty, kategorie i typy danych, które należy uwzględnić w przyszłym systemie.
- **Opracowanie modelu koncepcyjnego i logicznego:** najpierw rejestrowane są kluczowe podmioty i ich relacje, a następnie atrybuty, zasady i szczegółowa struktura.
- **Modelowanie fizyczne:** projektowana jest techniczna implementacja modelu: tabele, pola, relacje, ograniczenia, indeksy.
- **Tworzenie bazy danych:** ostatnim krokiem jest wdrożenie modelu fizycznego w wybranym systemie DBMS, przeprowadzenie testów i przygotowanie do pracy.



Rys. 4.3-5 Tworzenie baz danych i systemów zarządzania danymi dla procesów biznesowych rozpoczyna się od generowania wymagań i modelowania danych.

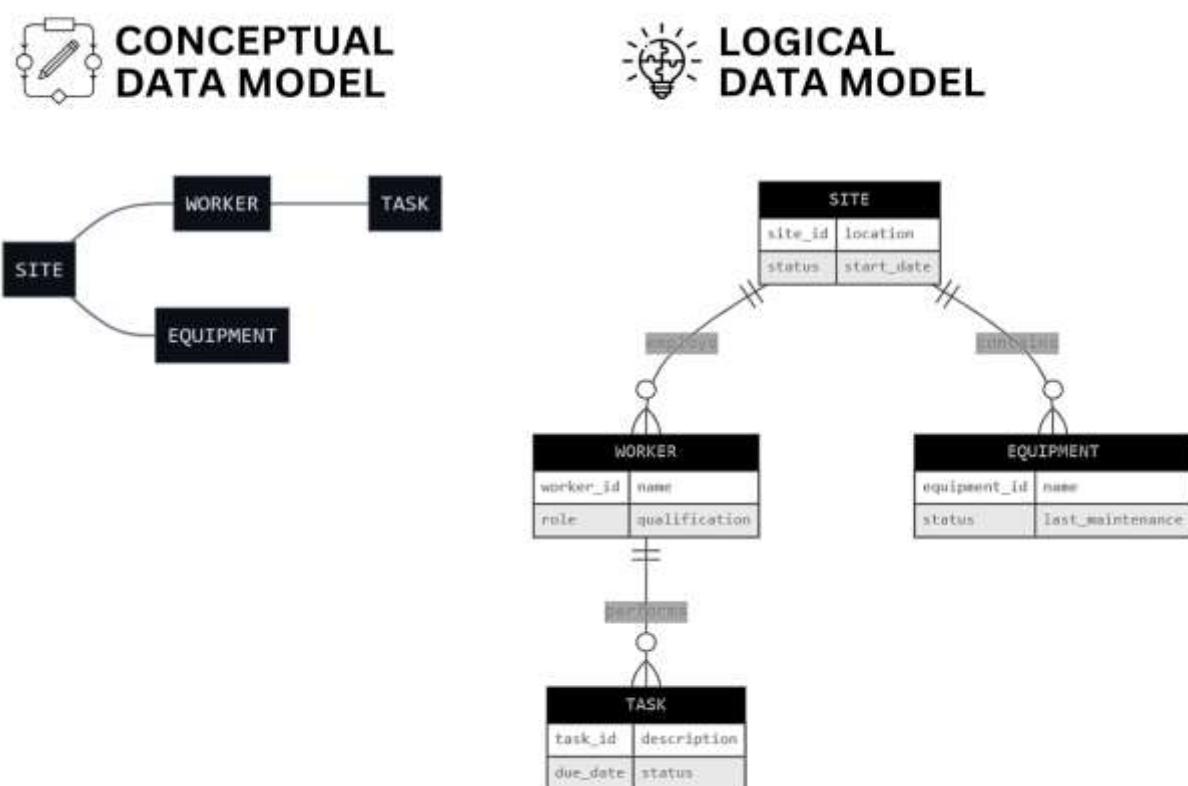
Prawidłowo zaprojektowane procesy modelowania danych pozwalają na przejrzysty przepływ informacji, co jest szczególnie ważne w złożonych projektach, takich jak zarządzanie projektem budowlanym lub placem budowy. Przyjrzyjmy się, w jaki sposób przejście od modelu koncepcyjnego do modelu logicznego, a następnie do modelu fizycznego może pomóc w usprawnieniu procesów.

Praktyczne modelowanie danych w kontekście budownictwa

Weźmy zadanie zarządzania placem budowy jako przykład modelowania danych i przekształćmy wymagania brygadzysty w ustrukturyzowany model logiczny. W oparciu o podstawowe potrzeby

zarządzania placem budowy, definiujemy kluczowe encje dla: placu budowy (SITE), pracowników (WORKER), sprzętu (EQUIPMENT), zadań (TASK) i wykorzystania sprzętu (EQUIPMENT_USAGE). Każda encja zawiera zestaw atrybutów, które odzwierciedlają ważne cechy. Na przykład w przypadku ZADANIA może to być opis zadania, termin wykonania, status, priorytet; w przypadku PRACOWNIKA może to być imię i nazwisko, rola w miejscu pracy, aktualne zatrudnienie itp.

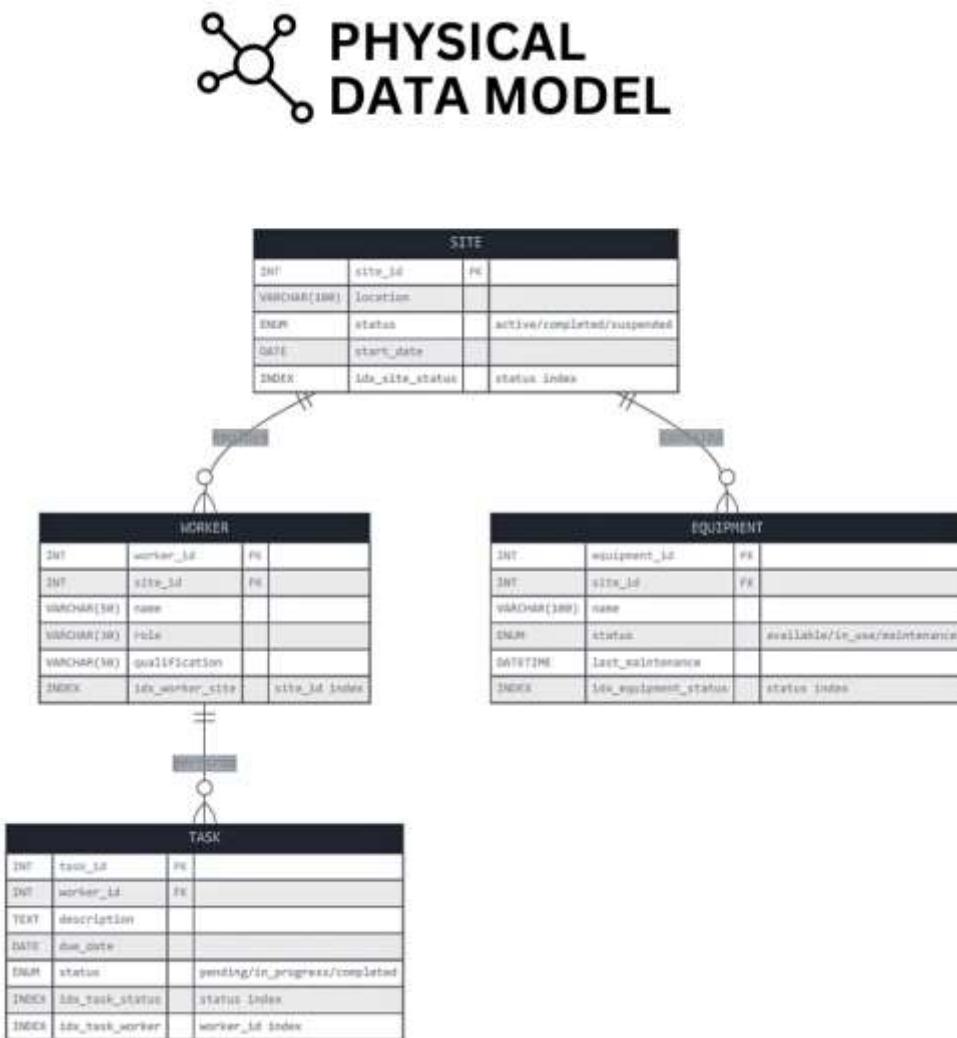
Model logiczny ustanawia relacje między tymi jednostkami, pokazując, w jaki sposób współdziałają one ze sobą w rzeczywistych procesach pracy (Rys. 4.3-6). Na przykład relacja między miejscem a pracownikami wskazuje, że wielu pracowników może pracować w jednym miejscu, podczas gdy relacja między pracownikami a zadaniami odzwierciedla, że jeden pracownik może wykonywać wiele zadań.



Rys. 4.3-6 Koncepcyjny i logiczny model danych wygenerowany na podstawie wymagań brygadisty dotyczących opisu procesów na placu budowy.

Po przejściu do modelu fizycznego dodawane są techniczne szczegóły implementacji: określone typy danych (VARCHAR, INT, DATE), klucze podstawowe i obce dla relacji między tabelami oraz indeksy w celu optymalizacji wydajności bazy danych (Rys. 4.3-7).

Na przykład dla statusów należy zdefiniować określone typy z możliwymi wartościami, a indeksy na kluczowych polach, takich jak status i worker_id, powinny zostać dodane w celu poprawy wydajności wyszukiwania. W ten sposób logiczny opis systemu staje się konkretnym planem implementacji bazy danych, gotowym do utworzenia i wdrożenia.



Rys. 4.3-7 Fizyczny model danych opisuje podmioty placu budowy za pomocą minimalnych wymaganych parametrów.

Model fizyczny często różni się od modelu logicznego. Średnio rozkład czasu modelowania jest następujący: około 50% czasu poświęca się na model koncepcyjny (zbieranie wymagań, omawianie procesów, identyfikacja podmiotów), 10% na model logiczny (określanie atrybutów i relacji) i 40% na model fizyczny (implementacja, testowanie, dostosowanie do DBMS).

Równowaga ta wynika z faktu, że etap koncepcyjny kładzie podwaliny pod strukturę danych, podczas gdy model logiczny określa jedynie relacje i atrybuty. Model fizyczny wymaga najwięcej zasobów, ponieważ to właśnie na tym etapie dane są wdrażane do konkretnych platform i narzędzi.

Tworzenie bazy danych przy użyciu LLM

Mając model danych i opis encji za pomocą parametrów, jesteśmy gotowi do tworzenia baz danych - magazynów, w których będziemy przechowywać informacje pochodzące z etapu strukturyzacji na temat konkretnych procesów.

Spróbujmy stworzyć przykład prostej, ale funkcjonalnej bazy danych z minimalną ilością kodu przy użyciu SQLite na przykładzie języka programowania Python. Relacyjne bazy danych zostały szczegółowo omówione w rozdziale "Strukturalne relacyjne bazy danych i język zapytań SQL".

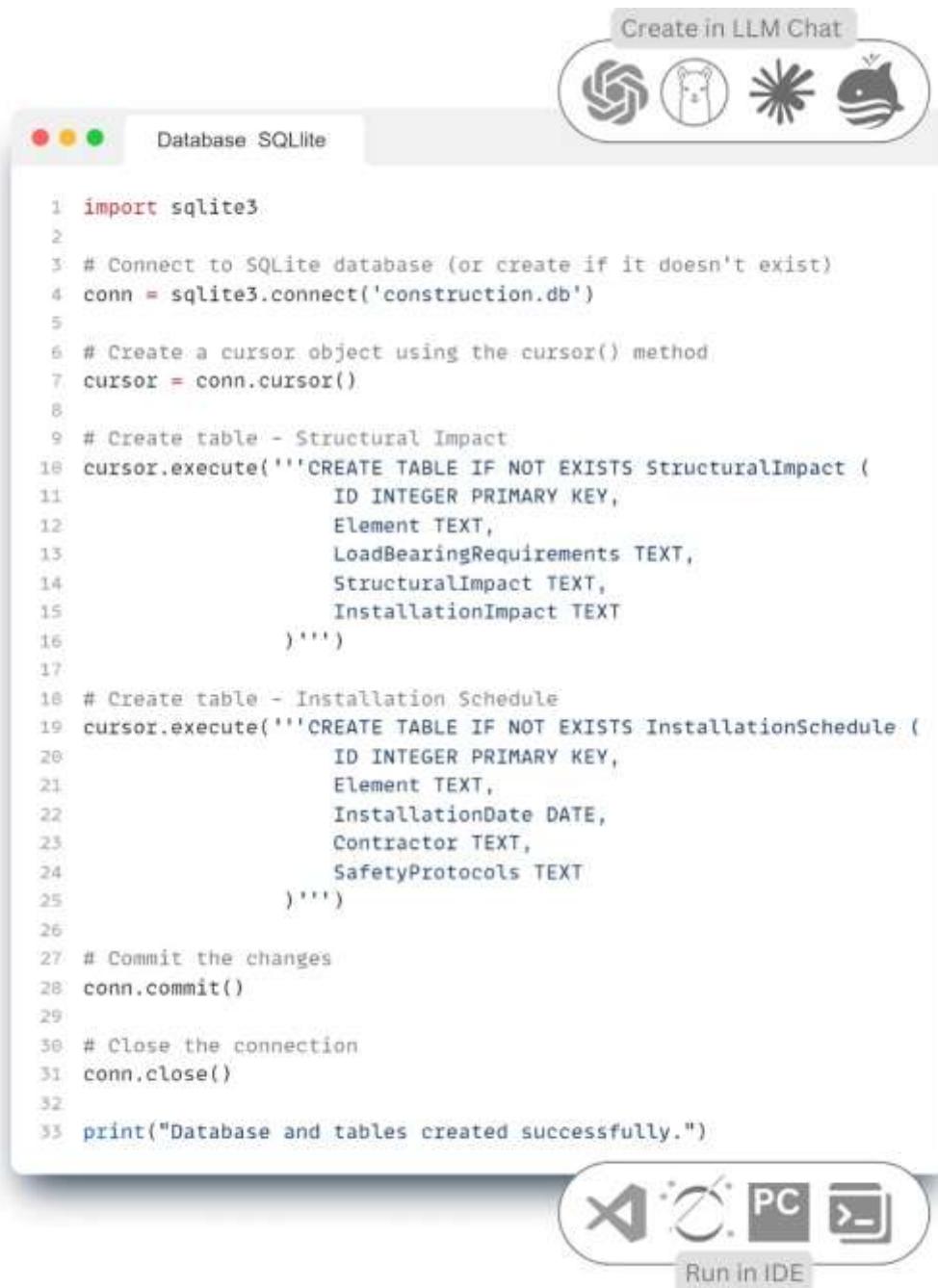
SQLite to lekka, open source'owa, wieloplatformowa baza danych, obsługująca podstawowe funkcje SQL. SQLite jest szeroko stosowany w CAD (BIM), ERP, EPM i jest obecny w prawie każdym smartfonie.

Poprośmy model LLM o utworzenie dla nas małej bazy danych SQLite w oparciu o model fizyczny utworzony powyżej (Rys. 4.3-4) dla systemów FEM (pływ strukturalny) i CPM (harmonogram montażu).

- 👉 Wyślij prośbę tekstową na czat LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN lub dowolny inny):

Utwórz bazy danych SQLite dla systemów "Constructive Impact" i "Installation Schedule" na podstawie rysunku w załączniku (wklej schemat do czatu jako obraz (rys. 4.3-4) ↴

2 Odpowiedź LLM :



```
1 import sqlite3
2
3 # Connect to SQLite database (or create if it doesn't exist)
4 conn = sqlite3.connect('construction.db')
5
6 # Create a cursor object using the cursor() method
7 cursor = conn.cursor()
8
9 # Create table - Structural Impact
10 cursor.execute('''CREATE TABLE IF NOT EXISTS StructuralImpact (
11         ID INTEGER PRIMARY KEY,
12         Element TEXT,
13         LoadBearingRequirements TEXT,
14         StructuralImpact TEXT,
15         InstallationImpact TEXT
16     )''' )
17
18 # Create table - Installation Schedule
19 cursor.execute('''CREATE TABLE IF NOT EXISTS InstallationSchedule (
20         ID INTEGER PRIMARY KEY,
21         Element TEXT,
22         InstallationDate DATE,
23         Contractor TEXT,
24         SafetyProtocols TEXT
25     )''' )
26
27 # Commit the changes
28 conn.commit()
29
30 # Close the connection
31 conn.close()
32
33 print("Database and tables created successfully.")
```

Create in LLM Chat

Run in IDE

Rys. 4.3-8 Wygenerowany kod tworzy bazy danych i tabele na podstawie przesłanego obrazu, do których możemy teraz załadować dane.

Ten kod bazy danych może być uruchamiany w środowisku Python z obsługą SQLite, które jest zwykle zawarte w standardowym środowisku Python. Można go uruchomić na dowolnej lokalnej maszynie Python zdolnej do uruchamiania skryptów i tworzenia plików. Po uruchomieniu i wykonaniu kod ten utworzy plik na dysku twardym - bazę danych SQLite o nazwie construction.db (podobną do pliku Excel z arkuszami roboczymi), która będzie zawierać tabele "Structural Impact" i "Installation Schedule".

Po utworzeniu nowej bazy danych **construction.db** w tych tabelach można dodawać dane za pomocą zapytań SQL lub importu, co pozwoli później rozpoczęć automatyczne przetwarzanie danych. Dane mogą być importowane do bazy danych SQLite z plików CSV, arkuszy kalkulacyjnych Excel lub eksportowane z innych baz danych i repozytoriów za pośrednictwem interfejsu API.

Aby ustanowić zrównoważone modelowanie danych i skuteczne procesy zarządzania bazami danych, firma potrzebuje jasno określonej strategii, a także koordynacji między zespołami technicznymi i biznesowymi. W przypadku różnych projektów i wielu źródeł danych często trudno jest zapewnić spójność, standaryzację i kontrolę jakości na wszystkich poziomach. Jednym z kluczowych rozwiązań może być stworzenie dedykowanego Centrum Doskonałości Modelowania Danych (CoE) w firmie.

Centrum doskonałości (CoE) w zakresie modelowania danych

Ponieważ dane stają się jednym z kluczowych aktywów strategicznych, firmy muszą zrobić coś więcej niż tylko prawidłowo gromadzić i przechowywać informacje - ważne jest, aby nauczyć się systematycznie zarządzać danymi. Centrum Doskonałości Klasifikacji i Modelowania Danych (CoE) jest jednostką strukturalną, która zapewnia spójność, jakość i wydajność całej obsługi danych w organizacji.

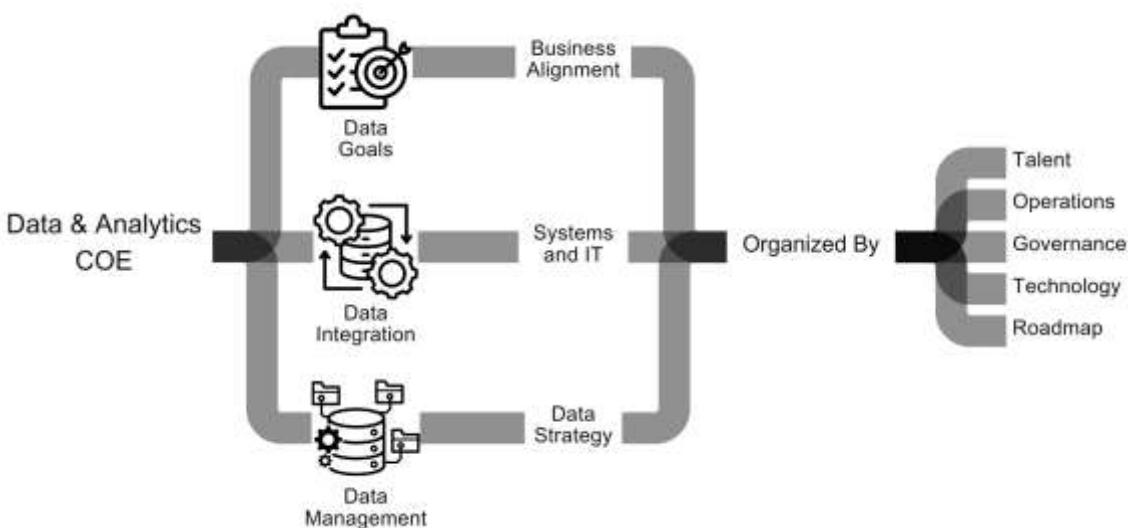
Centrum Doskonałości (CoE) jest rdzeniem wsparcia eksperckiego i fundamentem metodologicznym transformacji cyfrowej w firmie. Buduje kulturę opartą na danych i umożliwia organizacjom tworzenie procesów, które podejmują decyzje w oparciu o ustrukturyzowane, zweryfikowane i reprezentatywne dane, a nie intuicję lub informacje lokalne.

Centrum danych doskonałości jest zwykle tworzone z wielofunkcyjnych zespołów, które pracują zgodnie z zasadą "dwóch pizz". Zasada ta, zaproponowana przez Jeffa Bezosa, oznacza, że wielkość zespołu powinna być taka, aby można go było nakarmić dwiema pizzami, tj. nie więcej niż 6-10 osób. Takie podejście pozwala uniknąć nadmiernej biurokracji i zwiększa elastyczność pracy. Zespół CoE powinien obejmować pracowników o różnych umiejętnościach technicznych, od analityki danych i uczenia maszynowego po wiedzę specjalistyczną w określonych obszarach biznesowych. Dzięki dogłębnej wiedzy technicznej inżynierowie danych powinni nie tylko optymalizować procesy i modelować dane, ale także wspierać współpracowników, skracając czas wykonywania rutynowych zadań (rys. 4.3-9).

Tak jak w naturze odporność ekosystemu jest zapewniona przez bioróżnorodność, tak w świecie cyfrowym elastyczność i zdolność adaptacji są osiągane dzięki różnorodności podejść do obsługi danych. Różnorodność ta musi jednak opierać się na wspólnych zasadach i koncepcjach.

Centrum Doskonałości (CoE) można porównać do "warunków klimatycznych" ekosystemu leśnego, które określają, które rodzaje danych będą się rozwijać, a które zostaną automatycznie odrzucone. Tworząc

korzystny "klimat" dla wysokiej jakości danych, CoE ułatwia naturalną selekcję najlepszych praktyk i metodologii, które następnie stają się standardami dla organizacji.



Rys. 4.3-9 Centrum Doskonałości (CoE) w zakresie danych i analityki łączy wiedzę specjalistyczną na temat kluczowych aspektów zarządzania danymi, integracji i strategii.

Aby przyspieszyć cykle integracji i osiągnąć lepsze wyniki, CoE powinien zapewnić swoim członkom wystarczający stopień autonomii w podejmowaniu decyzji. Jest to szczególnie ważne w dynamicznym środowisku, w którym próby i błędy, ciągłe informacje zwrotne i częste wydania mogą przynieść znaczące korzyści. Autonomia ta jest jednak skuteczna tylko wtedy, gdy istnieje jasna komunikacja i wsparcie ze strony kierownictwa wyższego szczebla. Bez strategicznej wizji i koordynacji na najwyższym szczeblu, nawet najbardziej kompetentny zespół może napotkać bariery we wdrażaniu swoich inicjatyw.

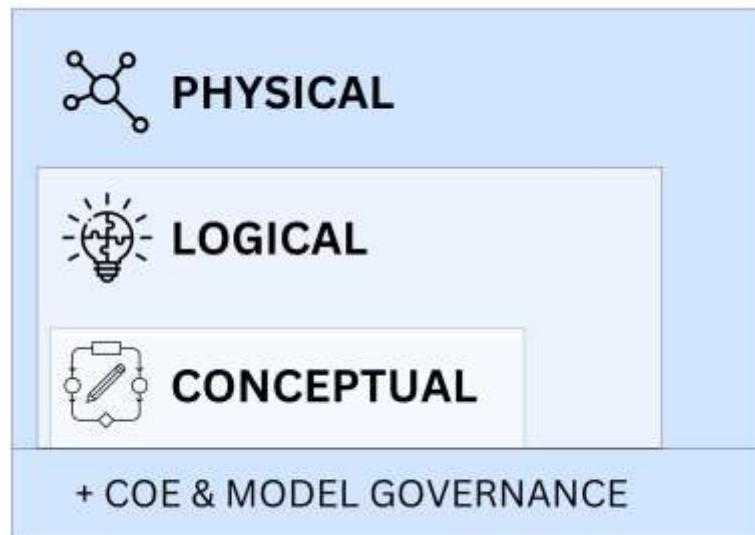
To COE lub kierownictwo wyższego szczebla firmy jest odpowiedzialne za zapewnienie, że podejście do modelowania danych nie ogranicza się do jednego lub dwóch projektów, ale jest osadzone w ogólnym systemie zarządzania informacjami i procesami biznesowymi.

Centrum wiedzy specjalistycznej (CoE), oprócz zadań związanych z modelowaniem danych i zarządzaniem danymi, jest odpowiedzialne za opracowywanie wspólnych standardów i podejść do wdrażania i obsługi infrastruktury danych. Ponadto wspiera kulturę ciągłego doskonalenia, optymalizacji procesów i efektywnego wykorzystania danych w organizacji (Rys. 4.3-10).

Systematyczne podejście do zarządzania danymi i modelami w ramach CoE można z grubsza podzielić na kilka kluczowych bloków:

- **Standaryzacja procesów i zarządzanie cyklem życia modelu:** CoE opracowuje i wdraża metodologie w celu ujednolicenia tworzenia modeli danych i zarządzania nimi. Obejmuje to: ustanowienie szablonów strukturalnych, metod kontroli jakości i systemów kontroli wersji w celu zapewnienia ciągłości danych na wszystkich etapach pracy.
- **Zarządzanie rolami i przypisywanie odpowiedzialności:** CoE definiuje kluczowe role w procesie modelowania danych. Każdy uczestnik projektu ma przypisane jasno określone role i obszary odpowiedzialności, co ułatwia pracę zespołową i zmniejsza ryzyko niespójności danych.

- **Kontrola jakości i audit:** skuteczne zarządzanie danymi budowlanymi wymaga ciągłego monitorowania ich jakości. Wdrażane są zautomatyzowane mechanizmy sprawdzania danych, identyfikowania błędów, brakujących atrybutów.
- **Zarządzanie metadanymi i architekturą informacji:** CoE odpowiada za stworzenie ujednoliconego systemu klasyfikacji i identyfikatorów, standardów nazewnictwa i opisu jednostek, co ma kluczowe znaczenie dla integracji między systemami.



Rys. 4.3-10 Modelowanie danych i zarządzanie jakością danych jest jednym z głównych wyzwań CoE

Centre of Excellence (CoE) for Data to nie tylko grupa ekspertów, ale systemowy mechanizm, który tworzy nową, opartą na danych kulturę i zapewnia jednolite podejście do pracy z danymi w całej firmie. Poprzez kompetentną integrację procesów modelowania z ogólnym systemem zarządzania informacjami, standaryzację, klasyfikację i kontrolę jakości danych, CoE pomaga firmom w ciągłym doskonaleniu ich produktów i procesów biznesowych, szybszym reagowaniu na zmiany rynkowe i podejmowaniu świadomych decyzji w oparciu o wiarygodne analizy.

Takie centra są szczególnie skuteczne w połączeniu z nowoczesnymi zasadami DataOps - pod adresem ruch, który zapewnia ciągłe dostarczanie, automatyzację i kontrolę jakości danych. Więcej o DataOps powiemy w części 8, w rozdziale "Nowoczesne technologie danych w branży budowlanej".

W kolejnych rozdziałach przejdziemy od strategii do praktyki - warunkowo "przekształćmy się" w centrum danych: przyjrzymy się kilku przykładom parametryzacji zadań, zbierania wymagań i automatycznego procesu walidacji.



ROZDZIAŁ 3.4.

USYSTEMATYZOWANIE WYMAGAŃ I WALIDACJA INFORMACJI

Gromadzenie i analiza wymagań: przekształca komunikację w ustrukturyzowane dane

Zbieranie wymagań i zarządzanie nimi jest pierwszym krokiem do zapewnienia jakości danych. Pomimo rozwoju narzędzi cyfrowych, większość wymagań jest nadal formułowana w formie nieustrukturyzowanej: poprzez listy, protokoły ze spotkań, rozmowy telefoniczne i dyskusje ustne. Taka forma komunikacji utrudnia automatyzację, walidację i ponowne wykorzystanie informacji. W tym rozdziale przyjrzymy się, jak przełożyć wymagania tekstowe na formalne struktury, zapewniając przejrzystość i systematyczność wymagań biznesowych.

Badanie firmy Gartner "Jakość danych: najlepsze praktyki w zakresie dokładnych informacji" podkreśla kluczowe znaczenie jakości danych dla udanych inicjatyw w zakresie danych i analityki [100]. Zauważają oni, że niska jakość danych kosztuje organizacje średnio co najmniej 12,9 miliona dolarów rocznie i że wiarygodne dane wysokiej jakości są niezbędne do stworzenia firmy opartej na danych.

Brak ustrukturyzowanych wymagań prowadzi do tego, że ten sam element (encja) i jego parametry mogą być przechowywane w różnych systemach w różnych wariantach. Nie tylko zmniejsza to wydajność procesów, ale także powoduje stratę czasu, powielanie informacji i konieczność ponownej walidacji danych przed ich wykorzystaniem. W rezultacie nawet pojedyncze pominięcie - zagubiony parametr lub pojedynczy nieprawidłowo opisany element - może spowolnić podejmowanie decyzji i spowodować nieefektywne wykorzystanie zasobów.

*Z braku gwoździa zgubiono podkowę.
Z braku podkowy zginął koń.
Z braku konia, jeździec został stracony.
Z braku jeźdźca, wiadomość została utracona.
Z braku wiadomości bitwa została przegrana.
Z braku bitwy, królestwo zostało utracone.
Wszystko przez brak gwoździa w podkowie.*

- Przysłowie [101]

Analiza i gromadzenie wymagań dotyczących procesu wypełniania i przechowywania danych rozpoczyna się od zidentyfikowania wszystkich interesariuszy. Tak jak przysłowiowa utrata pojedynczego gwoździa prowadzi do łańcucha krytycznych konsekwencji, tak w biznesie utrata pojedynczego interesariusza, przeoczone wymaganie lub utrata nawet pojedynczego parametru może znacząco wpłynąć nie tylko na pojedynczy proces biznesowy, ale na cały ekosystem projektu i organizacji jako całości. Dlatego kluczowe znaczenie ma identyfikacja nawet tych elementów, parametrów i ról, które na pierwszy rzut oka wydają się nieistotne, ale później mogą okazać się krytyczne dla zrównoważonego rozwoju biznesu.

Wyobraźmy sobie, że firma ma projekt, w którym klient zgłasza nową prośbę - "dodaj dodatkowe okno po północnej stronie budynku". Mały proces "prośba klienta o dodanie nowego okna do bieżącego projektu" obejmuje architekta, klienta, specjalistę CAD (BIM), kierownika budowy, kierownika logistyki, analityka ERP, inżyniera kontroli jakości, inżyniera bezpieczeństwa, kierownika kontroli i zarządcę nieruchomości.

Nawet niewielki proces może angażować dziesiątki różnych specjalistów. Każdy uczestnik procesu musi rozumieć wymagania specjalistów, z którymi jest połączony na poziomie danych.

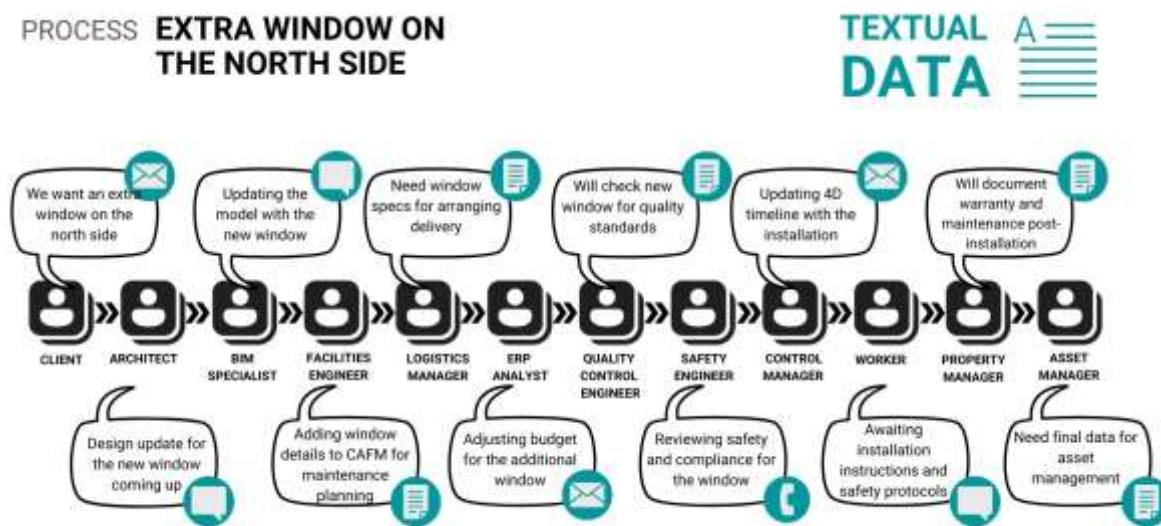
Na poziomie tekstu (rys. 4.4-1) komunikacja między klientem a specjalistami w łańcuchu procesów odbywa się w następujący sposób:

- ⦿ **Klient:** "Zdecydowaliśmy się dodać dodatkowe okno od strony północnej, aby uzyskać lepsze oświetlenie. Czy można to zrealizować?"
- ⦿ **Architekt:** "Jasne, poprawię projekt, aby uwzględnić nowe okno i prześlę zaktualizowane plany CAD (BIM)".
- ⦿ **Specjalista CAD (BIM):** "Otrzymałem nowy projekt. Aktualizuję model CAD (BIM) o dodatkowe okno i po uzgodnieniu z inżynierem FEM podaję dokładną lokalizację i wymiary nowego okna".
- ⦿ **Kierownik budowy:** "Otrzymaliśmy nowy projekt. Dostosowujemy daty instalacji 4D i informujemy wszystkich odpowiednich podwykonawców".
- ⦿ **Inżynier obiektów (CAFM):** "Wprowadzi dane 6D dotyczące nowego okna do systemu CAFM w celu przyszłego zarządzania obiektem i planowania konserwacji".
- ⦿ **Kierownik ds. logistyki:** "Potrzebuję wymiarów i wagi nowego okna, aby zorganizować jego dostawę na miejsce".
- ⦿ **ERP - analityk:** "Potrzebuję tabel zakresu i dokładnego typu okna dla aktualizacji budżetu 5D w naszym systemie ERP, aby odzwierciedlić koszt nowego okna w ogólnym oszacowaniu projektu".
- ⦿ **Inżynier ds. kontroli jakości:** "Gdy specyfikacje okien będą gotowe, upewnij się, że spełniają nasze standardy jakości i materiałów".
- ⦿ **Inżynier bezpieczeństwa:** "Będę oceniać aspekty bezpieczeństwa nowego okna, ze szczególnym uwzględnieniem zgodności i ewakuacji w ramach programu 8D".
- ⦿ **Controls Manager:** "W oparciu o dokładny zakres prac z ERP, zaktualizujemy naszą oś czasu 4D, aby odzwierciedlić instalację nowego okna i przechowywać nowe dane w systemie zarządzania treścią projektu".
- ⦿ **Pracownik (instalator):** "Potrzebne są instrukcje dotyczące instalacji, montażu i harmonogramu prac. Ponadto, czy wprowadzono jakieś specjalne zasady bezpieczeństwa, których muszę przestrzegać?".
- ⦿ **Zarządcy nieruchomości:** "Po zainstalowaniu udokumentuję informacje dotyczące gwarancji i konserwacji na potrzeby długoterminowego zarządzania".
- ⦿ **Asset Manager:** "Inżynierze ds. sprzętu, prosimy o przesłanie ostatecznych danych do śledzenia zasobów i zarządzania cyklem życia".
- ⦿ **Klient:** "Czekaj, może się spieszę i okno nie będzie potrzebne. Może powiniensem zrobić balkon".

W takich scenariuszach, które zdarzają się często, nawet niewielka zmiana powoduje reakcję łańcuchową między wieloma systemami i rolami. W takim przypadku prawie cała komunikacja w początkowej fazie odbywa się w formie tekstu: e-maile, czaty, protokoły ze spotkań (rys. 4.4-1).

W takim systemie komunikacji tekstu dla projektu budowlanego bardzo ważny jest system prawnego potwierdzania i rejestrowania wszystkich operacji wymiany danych i wszystkich podejmowanych decyzji.

Ma to na celu zapewnienie, że każda decyzja, instrukcja lub wprowadzona zmiana jest prawnie ważna i identyfikowalna, zmniejszając ryzyko przyszłych "nieporozumień"

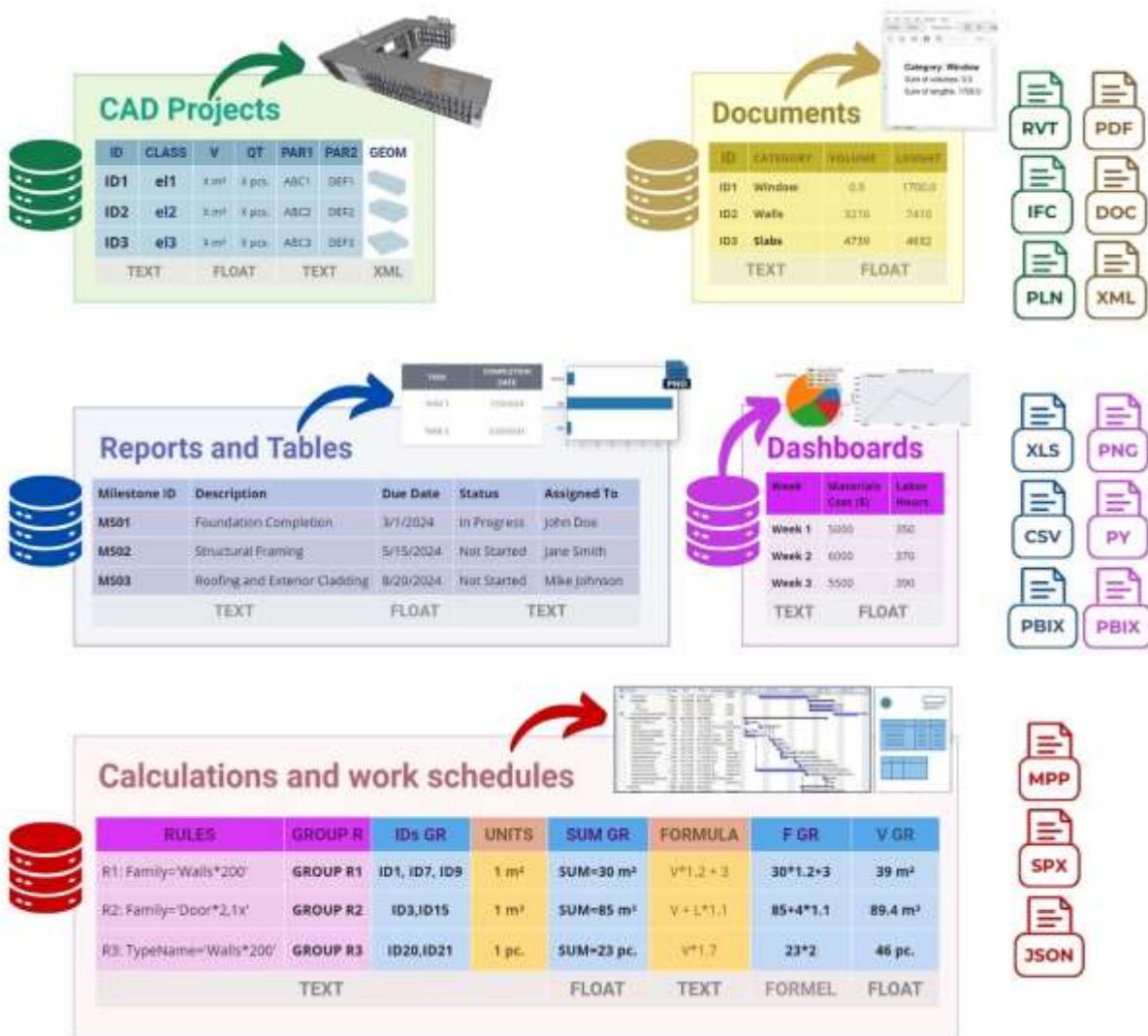


Rys. 4.4-1 Komunikacja między klientem a wykonawcą na początkowych etapach projektu często zawiera wieloformatowe dane tekstowe.

Brak kontroli prawnej i zatwierdzania decyzji w odpowiednich systemach projektu budowlanego może prowadzić do poważnych problemów dla wszystkich zaangażowanych stron. Każda decyzja, polecenie lub zmiana podjęta bez odpowiedniej dokumentacji i walidacji może prowadzić do sporów (i procesów sądowych).

Prawna konsolidacja wszystkich decyzji w komunikacji tekstuowej może być zapewniona tylko przez dużą liczbę podpisanych dokumentów, co spadnie na barki kierownictwa, które jest zobowiązane do rejestrowania wszystkich transakcji. W rezultacie, jeśli każdy uczestnik jest zobowiązany do podpisywania dokumentów dla każdego działania, system traci elastyczność i staje się biurokratycznym labiryntem. Brak potwierdzeń transakcji nie tylko opóźni realizację projektu, ale może również prowadzić do strat finansowych i pogorszenia relacji między uczestnikami, aż do problemów prawnych włącznie.

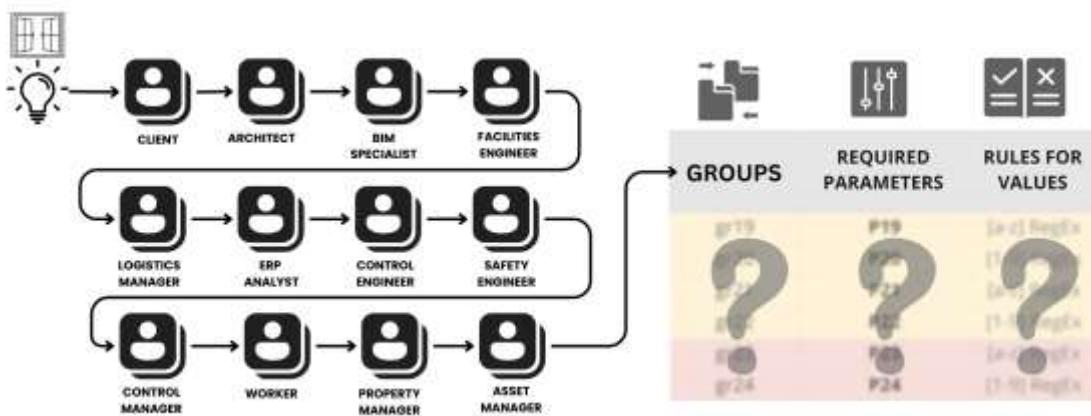
Taki proces zatwierdzania transakcji, który zwykle rozpoczyna się od dyskusji tekstowych, stopniowo przekształca się w wieloformatową wymianę dokumentów na kolejnych etapach (rys. 4.4-2), znacznie komplikując komunikację, która wcześniej odbywała się wyłącznie za pośrednictwem tekstu. Bez jasno zdefiniowanych wymagań automatyzacja takich procesów, które są wypełnione wieloformatowymi danymi i dużą liczbą wymagań tekstowych, staje się prawie niemożliwa.



Rys. 4.4-2 Każdy system w firmie budowlanej służy jako źródło istotnych prawnie dokumentów w różnych formatach.

Komunikacja tekstowa wymaga od każdego profesjonalisty zapoznania się z pełną korespondencją lub regularnego uczestniczenia we wszystkich spotkaniach, aby zrozumieć aktualny status projektu.

Aby przewyściżyć to ograniczenie, konieczne jest przejście od komunikacji tekstopisowej do ustrukturyzowanego modelu wymagań. Jest to możliwe tylko poprzez systematyczną analizę, wizualizację procesu i opis interakcji w postaci schematów blokowych i modeli danych (rys. 4.4-3). Podobnie jak w przypadku modelowania danych (rys. 4.3-7), przeszliśmy od poziomu kontekstowo-ideoowego do poziomu koncepcyjnego, dodając systemy i narzędzia używane przez uczestników oraz powiązania między nimi.



Rys. 4.4-3 Aby dowiedzieć się, jak zarządzać i zautomatyzować proces walidacji, konieczne jest zwizualizowanie procesów i ustrukturyzowanie wymagań.

Pierwszym krokiem w usystematyzowaniu wymagań i relacji jest wizualizacja wszystkich powiązań i relacji za pomocą koncepcyjnych schematów blokowych. Poziom koncepcyjny nie tylko ułatwi wszystkim uczestnikom procesu zrozumienie całego łańcucha procesów, ale także jasno pokaże, dlaczego i dla kogo dane (i wymagania) są potrzebne na każdym etapie procesu.

Schematy blokowe procesów i skuteczność ram koncepcyjnych

Aby wypełnić lukę między tradycyjnym a nowoczesnym podejściem do zarządzania danymi, firmy muszą świadomie przejść od fragmentarnych opisów tekstowych do ustrukturyzowanych reprezentacji procesów. Ewolucja danych - od glinianych tabliczek do cyfrowych ekosystemów - wymaga nowych narzędzi myślenia. Jednym z takich narzędzi jest modelowanie koncepcyjne z wykorzystaniem schematów blokowych. Tworzenie wizualnych diagramów - schematów blokowych, diagramów procesów, diagramów interakcji - pozwala uczestnikom projektu uświadomić sobie, w jaki sposób ich działania i decyzje wpływają na cały system decyzyjny.

Jeśli procesy wymagają nie tylko przechowywania danych, ale także ich analizy lub automatyzacji, należy zacząć zajmować się tematem tworzenia koncepcyjnej i wizualnej warstwy wymagań.

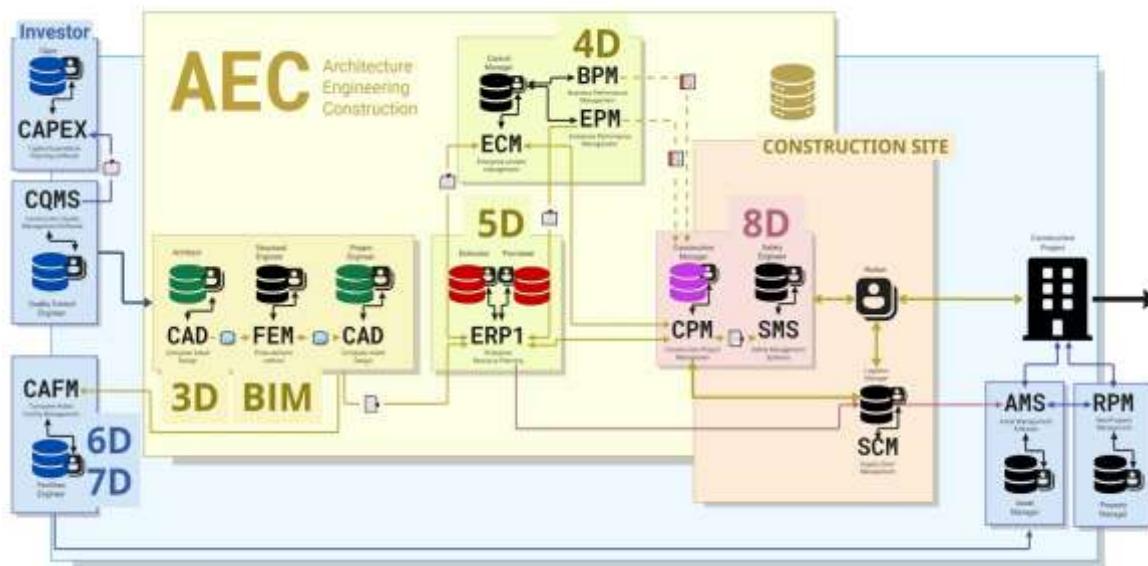
W naszym przykładzie (rys. 4.4-1) każdy specjalista może być częścią małego zespołu, ale także większego działu z maksymalnie kilkunastoma ekspertami pod kontrolą dyrektora generalnego. Każdy dział korzysta z wyspecjalizowanej bazy danych aplikacji (rys. 1.2-4 np. ERP, CAD, MEP, CDE, ECM, CPM itp.), która jest regularnie aktualizowana o napływaną informacje potrzebne do tworzenia dokumentów, rejestracji statusu prawnego decyzji i zarządzania procesami.

Proces transakcji jest podobny do pracy starożytnych menedżerów sprzed 4000 lat, kiedy to gliniane tabliczki i papirusy były używane do prawnego potwierdzania decyzji. Różnica między nowoczesnymi systemami a ich glinianymi i papierowymi poprzednikami polega na tym, że nowoczesne metody obejmują dodatkowo proces konwersji informacji tekstowych do postaci cyfrowej w celu dalszego automatycznego przetwarzania w innych systemach i narzędziach.

Stworzenie wizualizacji procesu w formie koncepcyjnych schematów blokowych pomoże opisać każdy krok i interakcje między różnymi rolami, czyniąc złożony przepływ pracy jasnym i prostym.

Wizualizacja procesów zapewnia, że logika procesu jest przejrzysta i dostępna dla wszystkich członków zespołu.

Ten sam proces komunikacyjny dodawania okna do projektu, który został opisany w formie tekstu, komunikatów (Rys. 4.4-1) i schematu blokowego, jest podobny do modelu koncepcyjnego, który omówiliśmy w rozdziale dotyczącym modelowania danych (Rys. 4.4-4).



Rys. 4.4-4 Schemat koncepcyjny przedstawia uczestników projektu jako użytkowników bazy danych, w której ich zapytania łączą różne systemy.

Chociaż diagramy koncepcyjne są ważnym krokiem, wiele firm ogranicza się do tego poziomu, wierząc, że diagram wizualny jest wystarczający do zrozumienia procesów. Stwarza to iluzję łatwości zarządzania: menedżerowie mogą łatwiej dostrzec duży obraz na takim schemacie blokowym i zobaczyć połączenia między uczestnikami i etapami. Takie schematy nie dają jednak jasnego wyobrażenia o tym, jakie dane są wymagane dla każdego uczestnika, w jakim formacie powinny być przesyłane oraz jakie parametry i atrybuty są obowiązkowe do realizacji automatyzacji. Koncepcyjny schemat blokowy bardziej przypomina mapę trasy: wskazuje, kto wchodzi w interakcję z kim, ale nie ujawnia, co jest przekazywane w tych interakcjach.

Nawet jeśli proces jest szczegółowo opisany na poziomie koncepcyjnym za pomocą schematów blokowych, nie gwarantuje to jego skuteczności. Wizualizacja często upraszcza pracę menedżerów, umożliwiając im łatwiejsze śledzenie procesu dzięki raportowaniu krok po kroku. Jednak dla inżynierów baz danych koncepcyjna reprezentacja może nie być wystarczająco jasna i może nie zapewniać jasnego zrozumienia sposobu wdrożenia procesu na poziomie parametrów i wymagań.

W miarę jak przechodzimy do bardziej złożonych ekosystemów danych, początkowe wdrożenie narzędzi koncepcyjnych i wizualnych staje się krytyczne dla zapewnienia, że procesy danych są nie tylko wydajne, ale także dostosowane do strategicznych celów organizacji. Aby w pełni przetłumaczyć ten proces dodawania okna (Rys. 4.4-1) na poziom wymagań dotyczących danych, musimy przejść poziom głębszej i przetłumaczyć koncepcyjną wizualizację procesu na logiczny i fizyczny poziom danych, wymaganych atrybutów i ich wartości granicznych.

Wymagania strukturalne i wyrażenia regularne RegEx

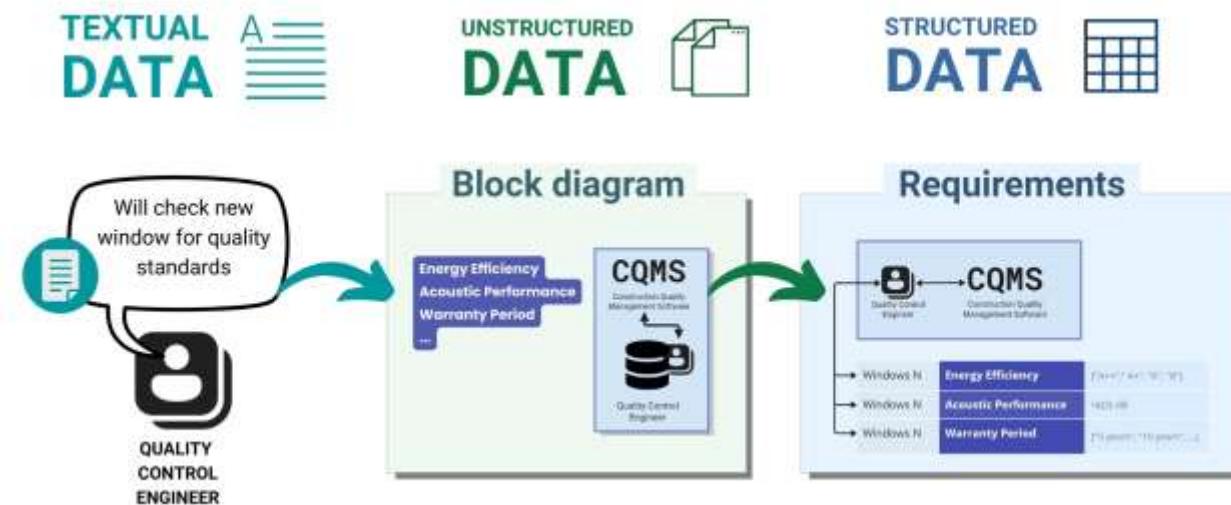
Do 80% danych tworzonych w firmach jest w formatach nieustrukturyzowanych lub częściowo ustrukturzowanych [52] - tekst, dokumenty, listy, pliki PDF, rozmowy. Takie dane (rys. 4.4-1) są trudne do analizy, weryfikacji, przesyłania między systemami i wykorzystania w automatyzacji.

Aby zapewnić łatwość zarządzania, przejrzystość i automatyczną walidację, konieczne jest przetłumaczenie tekstowych i częściowo ustrukturzowanych wymagań na dobrze zdefiniowane, ustrukturzowane formaty. Proces strukturyzacji dotyczy nie tylko danych (które szczegółowo omówiliśmy w pierwszych rozdziałach tej części książki), ale także samych wymagań, które uczestnicy projektu zwykle formułują w formie dowolnego tekstu przez cały cykl życia projektu, często nie myśląc o tym, że procesy te można zautomatyzować.

Tak jak już przekonwertowaliśmy dane z nieustrukturyzowanej formy tekstowej na formę ustrukturzowaną, tak w przepływie pracy nad wymaganiami przekonwertujemy wymagania tekstowe na ustrukturzowany format "warstwy logicznej i fizycznej".

W ramach przykładu dodawania okna (Rys. 4.4-1), następnym krokiem jest opisanie wymagań dotyczących danych w formie tabelarycznej. Ustrukturujemy informacje dla każdego systemu używanego przez uczestników projektu, określając kluczowe atrybuty i ich wartości graniczne.

Rozważmy na przykład jeden z takich systemów (rys. 4.4-5) - System Zarządzania Jakością Budowy (CQMS), który jest używany przez inżyniera kontroli jakości po stronie klienta. Z jego pomocą sprawdza on, czy nowy element projektu - w tym przypadku "nowe okno" - jest zgodny z ustalonymi normami i wymaganiami.



Rys. 4.4-5 Konwersja wymagań tekstowych do formatu tabeli z opisami atrybutów encji upraszcza ich zrozumienie przez innych specjalistów.

Jako przykład rozważmy kilka ważnych wymagań dla atrybutów podmiotów typu "systemy okienne" w systemie CQMS (rys. 4.4-6): efektywność energetyczna, parametry akustyczne i okres gwarancji. Każda kategoria obejmuje określone normy i specyfikacje, które należy wziąć pod uwagę podczas projektowania i instalacji systemów okiennych.



Rys. 4.4-6 Inżynier ds. kontroli jakości powinien sprawdzić nowe elementy typu okna pod kątem efektywności energetycznej, izolacji akustycznej i standardów gwarancyjnych.

Wymagania dotyczące danych, które inżynier kontroli jakości określa w formie tabeli, mają na przykład następujące wartości graniczne:

- **Klasa efektywności energetycznej okien** waha się od "A++", oznaczającej najwyższą efektywność, do "B", uważanej za minimalny akceptowalny poziom, a klasy te są reprezentowane przez listę dopuszczalnych wartości ["A++", "A+", "A", "A", "A", "B"].
- **Izolacyjność akustyczna okien**, mierzona w decybelach i pokazująca ich zdolność do redukcji

hałasu ulicznego, jest definiowana przez wyrażenie regularne \d{2}dB.

- **Atrybut "Okres gwarancji"** dla encji "Typ okna" zaczyna się od pięciu lat, ustawiając ten okres jako minimalny dozwolony przy wyborze produktu; wartości okresu gwarancji, takie jak ["5 lat", "10 lat" itp.] lub warunek logiczny ">5 (lat)" są również określone.

Zgodnie z zebranymi wymaganiami, w ramach ustalonych atrybutów, nowe elementy kategorii lub klasy okien z ocenami poniżej "B", takimi jak "C" lub "D", nie przejdą testu efektywności energetycznej. Izolacyjność akustyczna okien w danych lub dokumentach do Inżyniera QA powinna być oznaczona dwucyfrową liczbą, po której następuje postfiks "dB", np. "35 dB" lub "40 dB", a wartości spoza tego formatu, takie jak "9 D B" lub "100 decybeli" nie będą akceptowane (ponieważ nie przejdą wzorca dla ciągów RegEx). Okres gwarancji musi zaczynać się od minimum "5 lat", a okna z krótszymi okresami gwarancji, takimi jak "3 lata" lub "4 lata", nie będą spełniać wymagań opisanych przez Inżyniera Jakości w formacie tabeli.

Aby sprawdzić takie wartości parametrów atrybutów względem wartości granicznych z wymagań w procesie walidacji, używamy listy dozwolonych wartości ([A", "B", "C"], słowników ([A": "H1", "H2"; "B": W1, "W2"]], operacje logiczne (np. ">", "<", "<=", ">=", "==") dla wartości liczbowych) i wyrażenia regularne (dla wartości łańcuchowych i tekstowych, takich jak w atrybucie "Acoustic Performance"). Wyrażenia regularne są niezwykle ważnym narzędziem podczas pracy z wartościami łańcuchowymi.

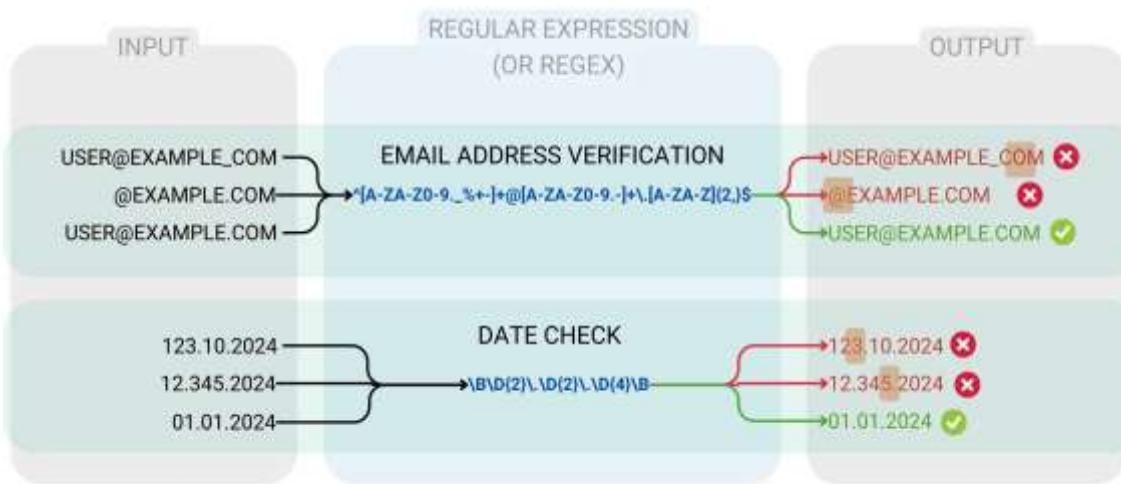
Wyrażenia regularne (RegEx) są używane w językach programowania, w tym w Pythonie (biblioteka Re), do wyszukiwania i modyfikowania ciągów znaków. Regex jest jak detektyw w świecie ciągów znaków, zdolny do precyzyjnego identyfikowania wzorców tekstowych w tekście.

W wyrażeniach regularnych litery są opisywane bezpośrednio za pomocą odpowiednich znaków alfabetu, podczas gdy liczby mogą być reprezentowane za pomocą znaku specjalnego \d, który odpowiada dowolnej cyfrze od 0 do 9. Nawiasy kwadratowe są używane do wskazania zakresu liter lub cyfr, np. [a-z] dla dowolnej małej litery alfabetu łacińskiego lub [0-9], co jest równoważne \d. W przypadku znaków nienumerycznych i nieliterowych używane są odpowiednio \D i \W.

Popularne przypadki użycia RegEx (Rys. 4.4-7):

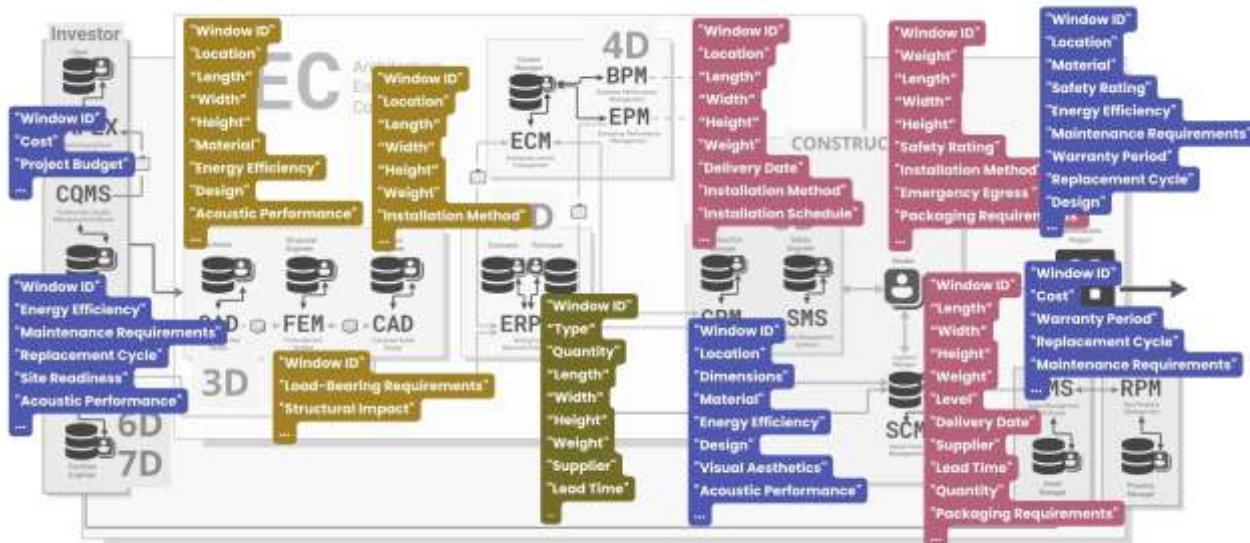
- **Weryfikacja adresu e-mail** : aby sprawdzić, czy ciąg znaków jest prawidłowym adresem e-mail, można użyć szablonu "^ [a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}\$".
- **Wyodrębnianie daty**: szablon "\b\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{4}\b" może być użyty do wyodrębnienia daty z tekstu w formacie DD.MM.RRRR.
- **Weryfikacja numerów telefonów** : aby zweryfikować numery telefonów w formacie +49(000)000-0000, wzorzec będzie wyglądał jak "\+\d{2}\(\d{3}\)\d{3}-\d{4}".

Przekładając wymagania inżyniera QA na format atrybutów i ich wartości brzegowych (rys. 4.4-6), przekształciliśmy je z ich oryginalnego formatu tekstowego (rozmowy, listy i dokumenty regulacyjne) w zorganizowaną i ustrukturyzowaną tabelę, umożliwiając w ten sposób automatyczne sprawdzanie i analizowanie wszelkich napływających danych (np. nowych elementów kategorii Okno). Obecność wymagań umożliwia automatyczne odrzucanie danych, które nie zostały sprawdzone, a także automatyczne przesyłanie sprawdzonych danych do systemów w celu dalszego przetwarzania.



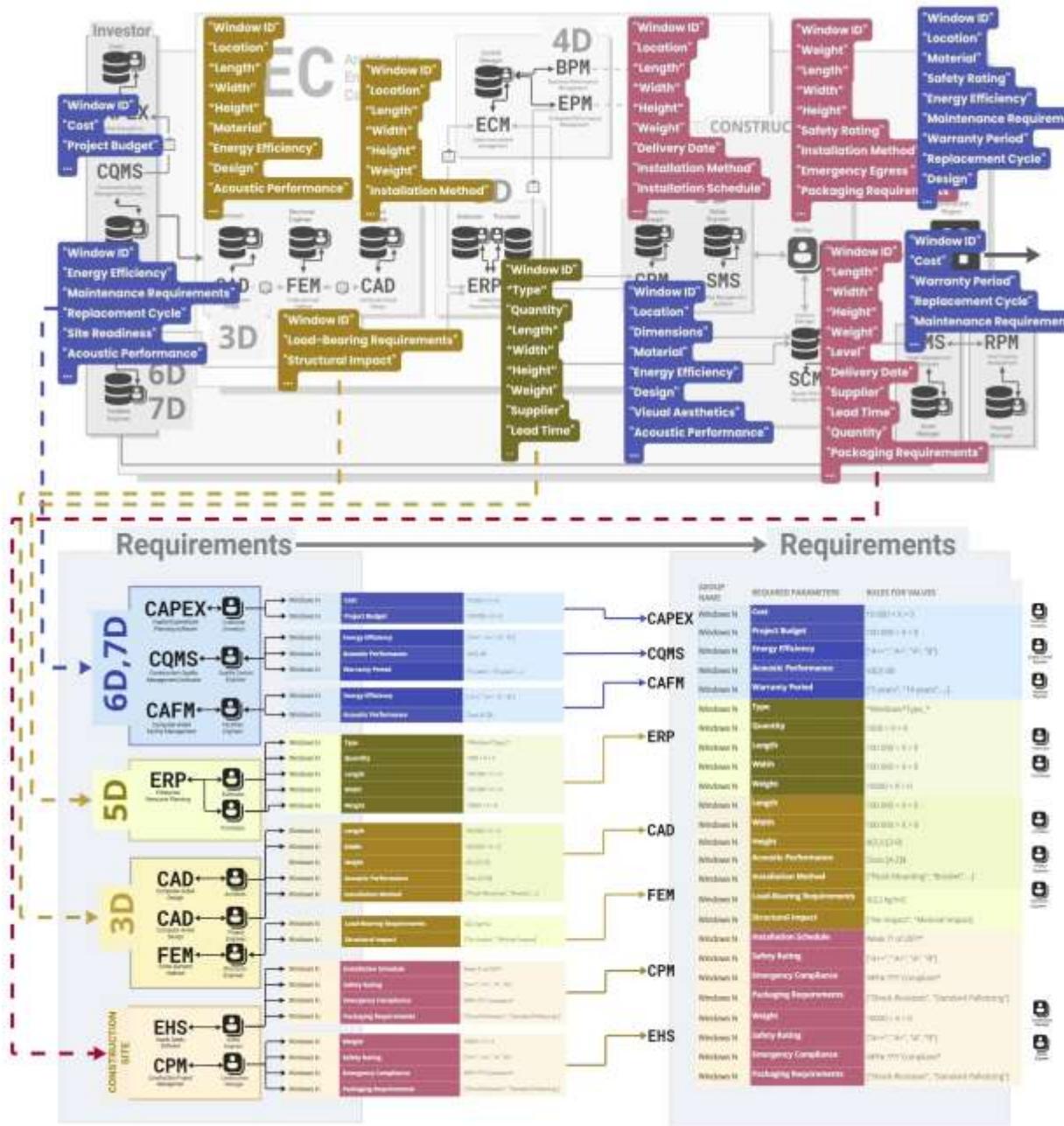
Rys. 4.4-7 Użycie wyrażeń regularnych jest niezwykle ważnym narzędziem w procesie walidacji danych tekstowych.

Teraz, przechodząc z konceptualnego do logicznego poziomu pracy z wymaganiami, przekonwertujemy wszystkie wymagania wszystkich specjalistów w naszym procesie instalacji nowego okna (rys. 4.4-4) na uporządkowaną listę w formacie atrybutów i dodamy te listy z niezbędnymi atrybutami do naszego schematu blokowego dla każdego specjalisty (rys. 4.4-8).



Rys. 4.4-8 Na poziomie procesu logicznego atrybuty obsługiwane przez każdego specjalistę są dodawane do odpowiednich systemów.

Dodając wszystkie atrybuty do jednej wspólnej tabeli procesu, przekształcamy informacje wcześniej prezentowane jako tekst i dialog na poziomie koncepcyjnym (rys. 4.4-1) w uporządkowaną i systematyczną formę tabel na poziomie fizycznym (rys. 4.4-9).



Rys. 4.4-9 Przekształcenie nieustrukturyzowanego dialogu specjalistów w ustrukturyzowane tabele pomaga zrozumieć wymagania na poziomie fizycznym.

Teraz wymagania dotyczące danych muszą zostać przekazane specjalistom tworzącym informacje dla określonych systemów. Na przykład, jeśli pracujesz w bazie danych CAD, przed rozpoczęciem modelowania elementów powinieneś zebrać wszystkie niezbędne parametry w oparciu o scenariusze końcowego wykorzystania danych. Zwykle zaczyna się to od fazy operacyjnej, a następnie placu budowy, działu logistyki, działu szacowania, działu obliczeń konstrukcyjnych i tak dalej. Dopiero po uwzględnieniu wymagań wszystkich tych obszarów można rozpocząć tworzenie danych - w oparciu o zebrane parametry. Pozwoli to w przyszłości zautomatyzować weryfikację i przesyłanie danych wzdłuż łańcucha.

Gdy nowe dane spełniają wymagania, są automatycznie integrowane z ekosystemem danych firmy, trafiając bezpośrednio do użytkowników i systemów, dla których są przeznaczone. Weryfikacja danych pod kątem atrybutów i ich wartości zapewnia, że informacje spełniają wymagane standardy jakości i są gotowe do wykorzystania w scenariuszach firmy.

Wymagania dotyczące danych zostały zdefiniowane, a teraz, przed rozpoczęciem weryfikacji, dane, które mają zostać zweryfikowane, muszą zostać utworzone, uzyskane lub zebrane, lub aktualny stan informacji w bazach danych musi zostać zarejestrowany w celu wykorzystania w procesie weryfikacji.

Gromadzenie danych dla procesu weryfikacji

Przed rozpoczęciem walidacji ważne jest, aby upewnić się, że dane są dostępne w formie odpowiedniej dla procesu walidacji. Nie oznacza to po prostu posiadania dostępnych informacji, ale ich przygotowanie: dane muszą zostać zebrane i przekształcone z nieustrukturyzowanych, luźno ustrukturyzowanych, tekstowych i geometrycznych formatów w ustrukturyzowaną formę. Proces ten został szczegółowo opisany w poprzednich rozdziałach, w których omówiono metody przekształcania różnych typów danych. W wyniku wszystkich przekształceń przychodzące dane przyjmują postać otwartych ustrukturyzowanych tabel (rys. 4.1-2, rys. 4.1-9, rys. 4.1-13).

Mając wymagania i ustrukturyzowane tabele z niezbędnymi parametrami i wartościami granicznymi (rys. 4.4-9), możemy rozpocząć walidację danych - albo jako pojedynczy zautomatyzowany proces (Pipeline), albo jako walidację krok po kroku każdego przychodzącego dokumentu.

Aby rozpocząć sprawdzanie, wymagane jest albo otrzymanie nowego pliku jako danych wejściowych, albo ustalenie bieżącego stanu danych - utworzenie migawki lub wyeksportowanie bieżących i przychodzących danych, albo skonfigurowanie połączenia z zewnętrzną lub wewnętrzną bazą danych. W rozważanym przykładzie taka migawka jest tworzona przez automatyczną konwersję danych CAD z do ustrukturyzowanego formatu zarejestrowanego, powiedzmy, o 23:00:00 w piątek, 29 marca 2024 r., po tym, jak wszyscy projektanci poszli do domu.



Rys. 4.4-10 Migawka bazy danych CAD (BIM) pokazująca aktualne informacje o atrybutie dla nowej encji klasy "Okno" w bieżącej wersji modelu projektu.

Dzięki narzędziom inżynierii odwrotnej omówionym w rozdziale "Przekształcanie danych CAD (BIM) w ustrukturyzowaną formę", informacje te z różnych narzędzi CAD (BIM) i edytorów mogą być zorganizowane

w oddzielne tabele (rys. 4.4-11) lub połączone w jedną wspólną tabelę łączącą różne sekcje projektu (rys. 9.1-10).

Taka tabela - baza danych wyświetla unikalne identyfikatory okien i drzwi (atrybut ID), nazwy typów (TypeName), wymiary (Szerokość, Długość), materiały (Materiał), a także wskaźniki charakterystyki energetycznej i akustycznej oraz inne cechy. Taka tabela wypełniona w programie CAD (BIM) jest zbierana przez inżyniera projektu z różnych działów i dokumentów, tworząc model informacyjny projektu.



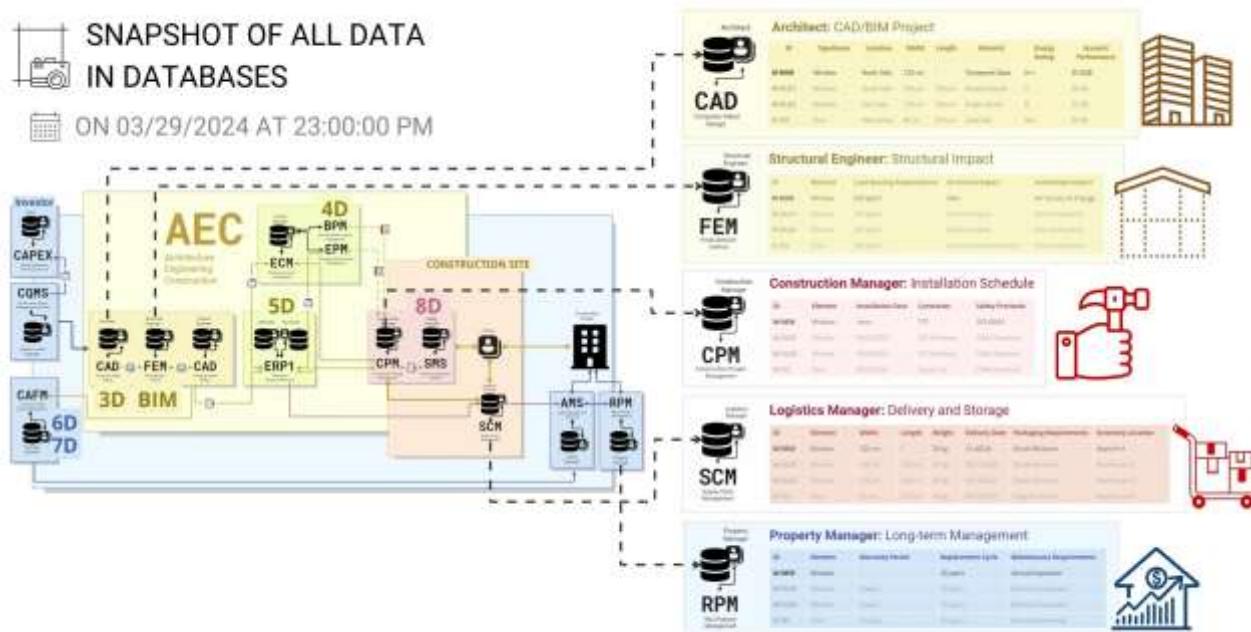
Rys. 4.4-11 Dane strukturalne z systemów CAD mogą być dwuwymiarową tabelą z kolumnami oznaczającymi atrybuty elementów.

Prawdziwe projekty CAD (BIM) zawierają dziesiątki lub setki tysięcy elementów (rys. 9.1-10). Elementy w formatach CAD są automatycznie kategoryzowane według typu i kategorii - od okien i drzwi po płyty, płyty i ściany. Unikalne identyfikatory (np. natywny identyfikator, który jest ustawiany automatycznie przez rozwiązanie CAD) lub atrybuty typu (nazwa typu, typ, rodzina) umożliwiają śledzenie tego samego obiektu w różnych systemach. Na przykład nowe okno na północnej ścianie budynku może być jednoznacznie identyfikowane za pomocą pojedynczego identyfikatora "W-NEW" we wszystkich odpowiednich systemach organizacji.

Podczas gdy nazwy i identyfikatory jednostek powinny być spójne we wszystkich systemach, zestaw atrybutów i wartości związanych z tymi jednostkami może się znacznie różnić w zależności od kontekstu użytkowania. Architekci, inżynierowie budowlani, specjalisci ds. budownictwa, logistyki i utrzymania nieruchomości postrzegają te same elementy na różne sposoby. Każdy z nich opiera się na własnych klasyfikatorach, standardach i celach: niektórzy patrzą na okno wyłącznie z estetycznego punktu widzenia, oceniając jego kształt i proporcje, podczas gdy inni patrzą na nie z inżynierijnego lub operacyjnego punktu widzenia, analizując przewodność cieplną, metodę instalacji, wagę lub wymagania konserwacyjne. Dlatego podczas modelowania danych i opisywania elementów ważne jest, aby wziąć pod uwagę wszechstronność ich zastosowania i zapewnić spójność danych, biorąc pod uwagę specyfikę branży.

Dla każdej roli w procesach firmy istnieją wyspecjalizowane bazy danych z własnym interfejsem użytkownika - od projektowania i obliczeń po logistykę, instalację i obsługę budynku (rys. 4.4-12). Każdy taki system jest zarządzany przez profesjonalny zespół specjalistów za pośrednictwem specjalnego interfejsu

użytkownika lub zapytań do bazy danych, gdzie za sumą wszystkich decyzji podejmowanych na podstawie wprowadzonych wartości na końcu łańcucha stoi kierownik systemu lub kierownik działu, który jest odpowiedzialny za ważność prawną i jakość wprowadzonych danych do swoich kontrahentów obsługujących inne systemy.



Rys. 4.4-12 Ta sama jednostka ma ten sam identyfikator w różnych systemach, ale różne atrybuty, które są ważne tylko w tym systemie.

Po zorganizowaniu zbierania ustrukturyzowanych wymagań i danych na poziomie logicznym i fizycznym, pozostało nam skonfigurować proces automatycznego sprawdzania danych z różnych dokumentów przychodzących i różnych systemów pod kątem wcześniej zebranych wymagań.

Weryfikacja danych i wyniki weryfikacji

Wszystkie nowe dane wprowadzane do systemu - czy to dokumenty, tabele czy wpisy do bazy danych od klienta, architekta, inżyniera, brygadzisty, logistyka czy zarządcy nieruchomości - muszą zostać zweryfikowane pod kątem sformułowanych wcześniej wymagań (rys. 4.4-9). Proces walidacji ma kluczowe znaczenie: wszelkie błędy w danych mogą prowadzić do nieprawidłowych obliczeń, opóźnienia harmonogramu, a nawet strat finansowych. Aby zminimalizować takie ryzyko, należy zorganizować systematyczną i powtarzaną, iteracyjną procedurę walidacji danych.

Aby zweryfikować nowe dane wprowadzane do systemu - nieuszkodzone, tekstowe lub geometryczne - należy je przekonwertować do formatu luźno ustrukturyzowanego lub ustrukturyzowanego. Proces walidacji musi następnie sprawdzić dane pod kątem pełnej listy wymaganych atrybutów i ich dozwolonych wartości.

Konwersja różnych typów danych: tekstu, obrazów, dokumentów PDF i mieszanych danych CAD (BIM) do postaci ustrukturyzowanej została szczegółowo omówiona w rozdziale "Konwersja danych do postaci ustrukturyzowanej".

Przykładem jest tabela uzyskana z projektu CAD (BIM) (rys. 4.4-11). Zawiera ona częściowo ustrukturyzowane dane geometryczne i ustrukturyzowane informacje o atrybutach elementów projektu (rys. 3.1-14) - np. element z klasy "Windows".

Aby przeprowadzić walidację, porównujemy wartości atrybutów (rys. 4.4-11) z referencyjnymi wartościami granicznymi, które zostały zdefiniowane przez ekspertów w formie wymagań (rys. 4.4-9). Ostateczna tabela porównawcza (rys. 4.4-13) wyjaśni, które wartości są akceptowalne, a które wymagają korekty, zanim dane będą mogły być wykorzystane poza aplikacjami CAD (BIM).



Rys. 4.4-13 Końcowa tabela walidacji podkreśla te wartości atrybutów dla nowej encji klasy "Windows", na które należy zwrócić uwagę.

Wdrażając podobne rozwiązanie przy użyciu biblioteki Pandas, którą opisaliśmy wcześniej w rozdziale "Pandas: Niezbędne narzędzie do analizy danych", sprawdzimy poprawność danych z pliku tabelarycznego wyodrębnionego z pliku CAD (BIM) (RVT, IFC, DWG, NWS, DGN) (Rys. 4.4-11) przy użyciu wymagań z innego tabelarycznego pliku wymagań (Rys. 4.4-9).

Aby uzyskać kod, musimy opisać w monicie dla LLM, że musimy załadować dane z pliku **raw_data.xlsx** (kompletny zestaw danych z bazy danych CAD (BIM)), **sprawdzić** je i zapisać wynik w nowym pliku **checked_data.xlsx** (rys. 4.4-13).

💡 Uzyskajmy kod przy użyciu LLM bez wspomniania o bibliotece Pandas:

Napisz kod, aby sprawdzić tabelę z pliku raw_data.xlsx i zweryfikować je przy użyciu następujących reguł walidacji: wartości kolumn "Szerokość" i "Długość" są większe od zera, "Ocena energetyczna" znajduje się na liście ["A++", "A+", "A", "B"], a "Charakterystyka akustyczna" jako zmienna, którą określmy później - z dodaną ostatnią kolumną walidacji i zapisz ostateczną tabelę w nowym pliku Excel checked_data.xlsx ↴

- Odpowiedź LLM będzie opisywać krótki przykład kodu Pythona, który może być udoskonalany i rozszerzany przez kolejne podpowiedzi :

```

1 import pandas as pd
2
3 # Load raw data from Excel file
4 df_data = pd.read_excel('raw_data.xlsx', sheet_name=0)
5
6 # Adding a 'Requirements Met' column that checks all conditions
7 df_data['Requirements Met'] =
8     (df_data['Width'] > 0) &
9     (df_data['Length'] > 0) &
10    (df_data['Energy Rating'].isin(['A++', 'A+', 'A', 'B'])) &
11    (df_data['Acoustic Performance'] ≥ min_acoustic_performance)
12 )
13
14 # Save the results to a new Excel file
15 df_data.to_excel('checked_data.xlsx', index=False)

```

Rys. 4.4-14 Kod wygenerowany przez model LLM- sprawdza przekonwertowany projekt CAD (BIM) pod kątem wymagań atrybutów w postaci wartości granicznych.

Kod wygenerowany przez model językowy LLM, może być używany w dowolnym popularnym IDE lub narzędziu online: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse z wtyczką PyDev, Thonny, Wing IDE, IntelliJ IDEA z wtyczką Python, JupyterLab lub popularnych narzędziach online Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Wykonanie kodu (rys. 4.4-14) pokaże, że "elementy encji" W-OLD1, W-OLD2, D-122 (i inne elementy) z bazy danych CAD (BIM) spełniają wymagania atrybutów: szerokość i długość są większe od zera, a klasa efektywności energetycznej jest jedną z wartości listy "A++", "A", "B", "C" (rys. 4.4-15).

Potrzebny i niedawno dodany element W-NEW, który odpowiada za nową klasę elementów "Window" po stronie północnej, jest niezgodny (atrybut "Requirments Met"), ponieważ jego długość wynosi zero (wartość "0,0" jest uważana za niedopuszczalną przez naszą regułę "Width">>0) i nie określa klasy efektywności

energetycznej.

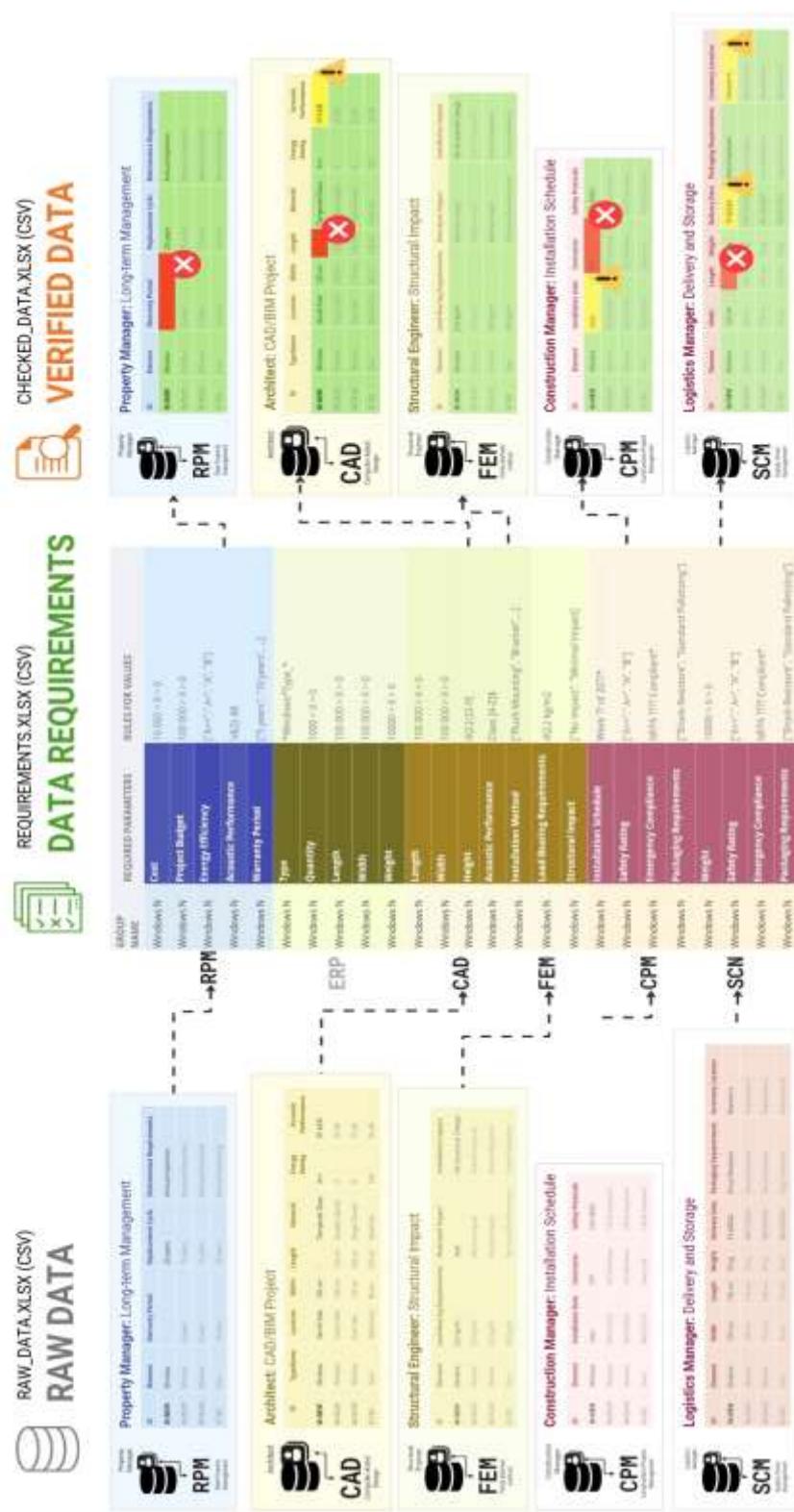
ID	TypeName	Location	Width	Length	Material	Energy Rating	Acoustic Performance	Requirements Met	
0	W-NEW	Window	North Side	120	0.0	Tempered Glass		35	False
1	W-OLD1	Window	South Side	100	140.0	Double Glazed	A++	30	True
2	W-OLD2	Window	East Side	110	160.0	Single Glazed	B	25	True
3	D-122	Door	Main Entry	90	210.0	Solid Oak	B	30	True

Rys. 4.4-15 Checking identyfikuje podmioty, które nie przeszły procesu weryfikacji i dodaje do wyników nowy atrybut z wartościami "False" lub "True".

Podobnie sprawdzamy spójność wszystkich elementów projektu (encji) i wymaganych atrybutów dla każdego z systemów, tabel lub baz danych we wszystkich danych, które otrzymujemy od różnych specjalistów (Rys. 4.4-1) podczas procesu dodawania okna do projektu.

W końcowej tabeli wygodnie jest wyróżnić wyniki kontroli kolorem w celu wizualizacji: zielony jest używany do oznaczania atrybutów, które pomyślnie przeszły kontrolę, żółty - wartości z niekrytycznymi odchyleniami, a czerwony - krytyczne rozbieżności (rys. 4.4-16).

W wyniku walidacji (Rys. 4.4-16) otrzymujemy listę zaufanych i zwalidowanych elementów wraz z ich identyfikatorami, które zostały zweryfikowane pod kątem spełnienia wymagań atrybutów. Zatwierdzone elementy dają pewność, że elementy te spełniają określone standardy i specyfikacje dla wszystkich systemów zaangażowanych w proces dodawania elementów klasy Window lub dowolnej innej klasy (więcej na temat automatyzacji walidacji danych i tworzenia zautomatyzowanego procesu ETL omówimy w rozdziale "Automatyzacja ETL i walidacji danych").



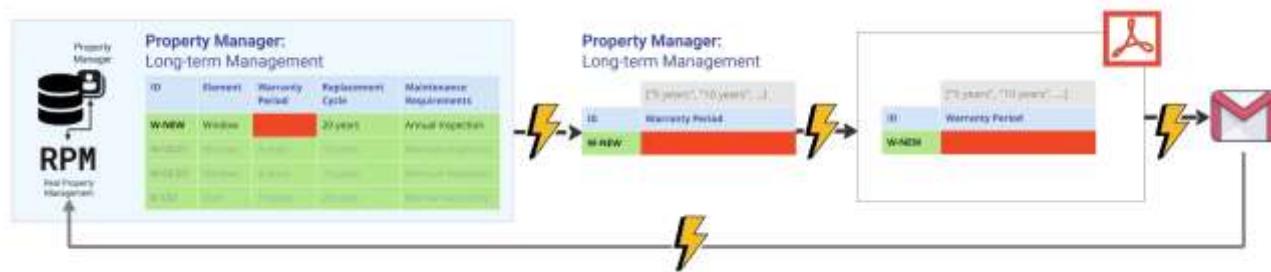
Rys. 4.4-16 Wynik kontroli przeprowadzonej dla wszystkich systemów pozwala określić, które dane nie spełniają wymagań firmy.

Podmioty, które pomyślnie przeszły walidację, zazwyczaj nie wymagają większej uwagi. Bez przeszkód przechodzą do kolejnych etapów przetwarzania i integracji z innymi systemami. W przeciwieństwie do elementów "wysokiej jakości", najbardziej interesujące są elementy, które nie przeszły walidacji. Informacje o takich odchyleniach są krytyczne: powinny być przekazywane nie tylko w raportach tabelarycznych, ale także za pomocą różnych narzędzi wizualizacyjnych. Graficzne przedstawienie wyników weryfikacji pomaga szybko ocenić ogólny stan jakości danych, zidentyfikować obszary problemowe i niezwłocznie podjąć działania naprawcze lub korygujące.

Wizualizacja wyników weryfikacji

Wizualizacja jest niezbędnym narzędziem do interpretacji wyników kontroli. Oprócz zwykłych tabel podsumowujących, może zawierać panele informacyjne, diagramy i automatycznie generowane dokumenty PDF, które grupują elementy projektu według ich statusu kontroli. Kodowanie kolorami może odgrywać tutaj rolę pomocniczą: zielony może wskazywać elementy, które zostały pomyślnie zweryfikowane, żółty może wskazywać elementy, które wymagają dodatkowej uwagi, a czerwony może wskazywać elementy, w których występują błędy krytyczne lub brakuje kluczowych danych.

W naszym przykładzie (rys. 4.4-1) analizujemy dane z każdego systemu krok po kroku: od CAD (BIM) i zarządzania nieruchomościami po logistykę i harmonogramy instalacji (rys. 4.4-16). Po przeprowadzeniu audytu dla każdego specjalisty automatycznie generowane są indywidualne alerty lub raporty, np. w formacie PDF (rys. 4.4-17). Jeśli dane są prawidłowe, specjalista otrzymuje krótką wiadomość: "Dziękujemy za współpracę". W przypadku stwierdzenia rozbieżności wysyłany jest szczegółowy raport o następującej treści: "Ten dokument zawiera listę elementów, ich identyfikatorów, atrybutów i wartości, które nie zostały sprawdzone pod kątem zgodności".



Rys. 4.4-17 Walidacja i automatyczne generowanie raportów przyspiesza proces znajdowania i rozumienia braków danych dla specjalisty, który je tworzy.

Dzięki zautomatyzowanemu procesowi walidacji - gdy tylko zostanie wykryty błąd lub luka w danych, natychmiast wysyłane jest powiadomienie w formie wiadomości czatu, e-maila lub dokumentu PDF do osoby odpowiedzialnej za tworzenie lub przetwarzanie odpowiednich encji i ich atrybutów (rys. 4.4-18), z listą elementów i opisów atrybutów, które nie zostały zwalidowane.



Rys. 4.4-18 Automatyczne raporty z inspekcji ułatwiają zrozumienie błędów i przyspieszają pracę nad uzupełnianiem danych projektu.

Na przykład, jeśli dokument dotrze do systemu zarządzania nieruchomościami (po strukturyzacji) z nieprawidłowo wypełnionym atrybutem "Okres gwarancji", zarządcą nieruchomości otrzyma alert z listą atrybutów, które należy sprawdzić i poprawić.

Podobnie, wszelkie niedociągnięcia w harmonogramie instalacji lub danych logistycznych spowodują wygenerowanie automatycznego raportu i, na przykład, wysłanie powiadomienia na czacie lub wiadomości e-mail z wynikami kontroli do odpowiedniego specjalisty.

Oprócz dokumentów PDF i wykresów z wynikami, możliwe jest tworzenie pulpitów nawigacyjnych i interaktywnych modeli 3D (Rys. 7.1-6, Rys. 7.2-12) podkreślających elementy z brakującymi atrybutami, umożliwiając użytkownikom wizualne wykorzystanie geometrii 3D elementów do filtrowania i oceny jakości i kompletności tych elementów w projekcie.

Wizualizacja wyników inspekcji w postaci automatycznie generowanych dokumentów, wykresów lub pulpitów znacznie upraszcza interpretację danych i ułatwia skuteczną komunikację między uczestnikami projektu.

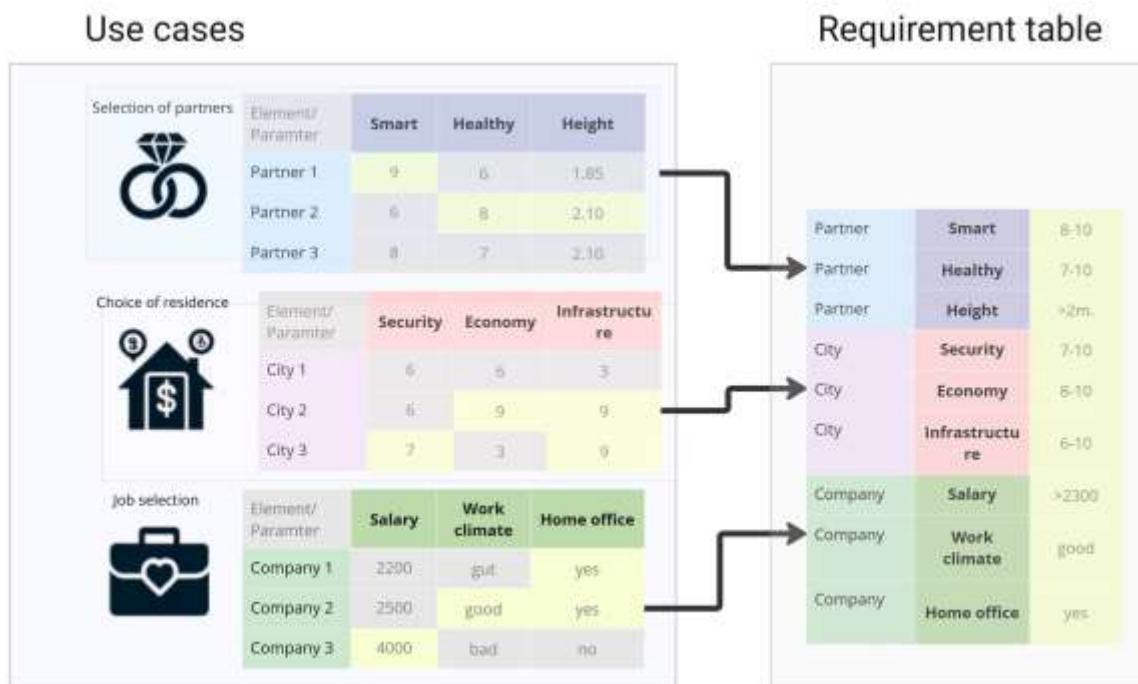
Proces automatycznego sprawdzania danych z różnych systemów i źródeł informacji można porównać do świadomego podejmowania decyzji w życiu codziennym. Tak jak firmy z branży budowlanej biorą pod uwagę wiele zmiennych - od wiarygodności danych wejściowych po ich wpływ na czas, koszt i jakość realizacji projektu - tak samo przy podejmowaniu ważnych decyzji, na przykład przy wyborze miejsca zamieszkania, człowiek rozważa cały szereg czynników: dostępność transportową, infrastrukturę, koszty, bezpieczeństwo, jakość życia. Wszystkie te czynniki tworzą system kryteriów, które stanowią podstawę ostatecznych decyzji składających się na nasze życie.

Porównanie kontroli jakości danych z potrzebami ludzkiego życia

Pomimo ciągłego rozwoju metod i narzędzi kontroli jakości danych, fundamentalna zasada zgodności informacji pozostaje niezmienna. Zasada ta stanowi fundament dojrzałego systemu zarządzania, zarówno w biznesie, jak i w życiu codziennym.

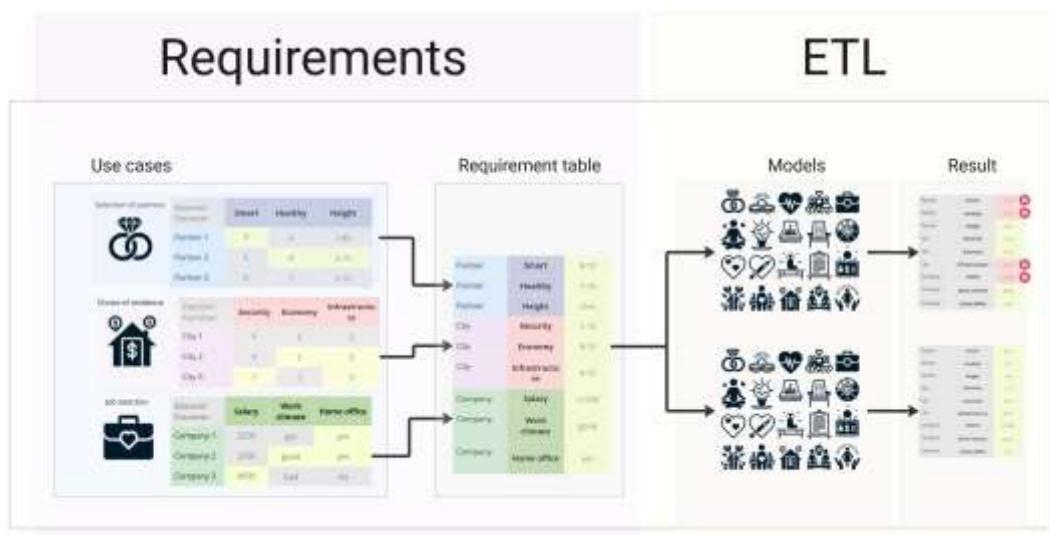
Proces iteracyjnej walidacji danych jest bardzo podobny do procesu podejmowania decyzji, z którym każdy ma do czynienia na co dzień. W obu przypadkach polegamy na doświadczeniu, danych i nowych informacjach, które stają się dostępne. Coraz więcej decyzji życiowych i zawodowych - od strategicznych po codzienne - podejmowanych jest w oparciu o dane.

Na przykład, wybierając miejsce zamieszkania lub partnera życiowego, intuicyjnie tworzymy w naszych umysłach tabelę kryteriów i cech, według których porównujemy alternatywy (rys. 4.4-19). Cechy te - czy to osobiste cechy osoby, czy parametry nieruchomości - reprezentują atrybuty, które wpływają na ostateczną decyzję.



Rys. 4.4-19 Wybór miejsca zamieszkania, zatrudnienia lub partnerstwa opiera się na indywidualnych wymaganiach dotyczących atrybutów.

Wykorzystanie ustrukturyzowanych danych i sformalizowanego podejścia do opisywania wymagań (Rys. 4.4-20) przyczynia się do bardziej świadomych i przemyślanych wyborów zarówno w życiu zawodowym, jak i osobistym.

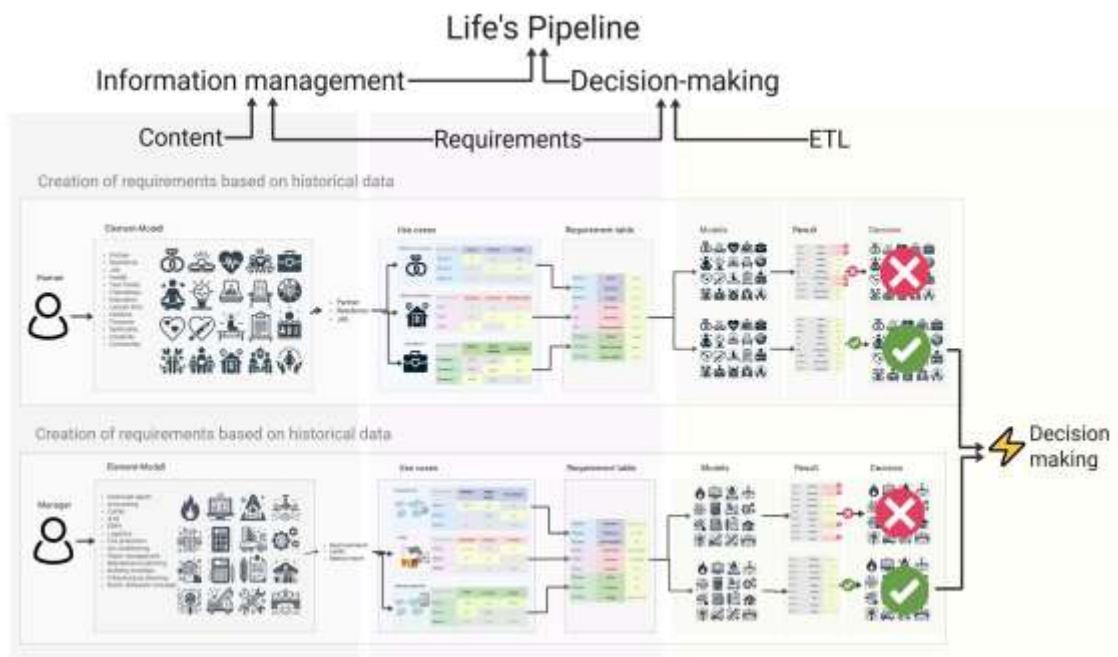


Rys. 4.4-20 Formalizacja wymagań pomaga usystematyzować postrzeganie życia i decyzji biznesowych.

Podejście do podejmowania decyzji oparte na danych nie jest wyłącznie narzędziem biznesowym. Jest ono również płynnie zintegrowane z codziennym życiem, podążając za wspólnymi krokami przetwarzania danych (rys. 4.4-21) podobnymi do procesu ETL (Extract, Transform, Load), który omówiliśmy już na początku tej części podczas strukturyzacji danych i który szczegółowo przeanalizujemy w kontekście automatyzacji zadań w siódmej części książki:

- **Dane jako podstawa (fragment):** W każdej dziedzinie - czy to w pracy, czy w życiu osobistym - gromadzimy informacje. W biznesie mogą to być raporty, liczby, dane rynkowe; w życiu osobistym mogą to być osobiste doświadczenia, porady od bliskich, informacje zwrotne, obserwacje.
- **Kryteria oceny (Transform):** zebrane informacje są interpretowane na podstawie wcześniej zdefiniowanych kryteriów. W pracy są to wskaźniki wydajności (KPI), ograniczenia budżetowe i normy; w życiu osobistym parametry takie jak cena, wygoda, niezawodność, charyzma itp.
- **Przewidywanie i analiza ryzyka (Load):** ostatni etap obejmuje podejmowanie decyzji w oparciu o analizę przekształconych danych i porównanie możliwych konsekwencji. Jest to podobne do procesów biznesowych, w których dane przechodzą przez logikę biznesową i filtr ryzyka.

Decyzje, które podejmujemy - od trywialnych preferencji, takich jak to, co zjeść na śniadanie, po ważne wydarzenia życiowe, takie jak wybór kariery lub partnera życiowego - są z natury wynikiem przetwarzania i oceny danych.



Rys. 4.4-21 Biznes i ogólnie życie to seria decyzji opartych na danych, w których jakość danych wykorzystywanych do podejmowania decyzji jest kluczowym czynnikiem.

Wszystko w naszym życiu jest ze sobą powiązane i tak jak organizmy żywe, w tym ludzie, podążają za prawami natury, ewoluując i dostosowując się do zmieniających się warunków, tak procesy ludzkie, w tym sposób gromadzenia i analizowania danych, odzwierciedlają te naturalne zasady. Ścisły związek między naturą a działalnością człowieka potwierdza nie tylko naszą zależność od natury, ale także nasze pragnienie stosowania praw doskonalonych przez miliony lat ewolucji w celu tworzenia architektur danych, procesów i systemów do podejmowania decyzji.

Nowe technologie, zwłaszcza w budownictwie, są doskonałym przykładem tego, jak ludzkość raz po raz inspiruje się naturą, aby tworzyć lepsze, bardziej zrównoważone i wydajne rozwiązania.

Kolejne kroki: przekształcanie danych w dokładne obliczenia i plany

W tej części przyjrzelismy się sposobom przekształcania nieustrukturyzowanych danych w ustrukturowany format, opracowywania modeli danych i organizowania procesów sprawdzania jakości informacji w projektach budowlanych. Zarządzanie danymi, ich standaryzacja i klasyfikacja to fundamentalny proces, który wymaga systematycznego podejścia i jasnego zrozumienia wymagań biznesowych. Techniki i narzędzia omówione w tej części pozwalają na niezawodną integrację między różnymi systemami w całym cyklu życia obiektu.

Podsumowując tę część, podkreślmy główne praktyczne kroki, które pomogą ci zastosować omówione podejścia w codziennych zadaniach:

- Zaczniij od usystematyzowania wymagań
 - Utwórz rejestr atrybutów i parametrów dla kluczowych elementów projektów i procesów.
 - Udokumentuj wartości graniczne dla każdego atrybutu

- Wizualizacja procesów i relacji między klasami, systemami i atrybutami za pomocą schematów blokowych (np. w Miro, Canva, Visio).

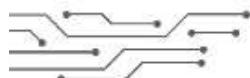
■ Automatyzacja konwersji danych

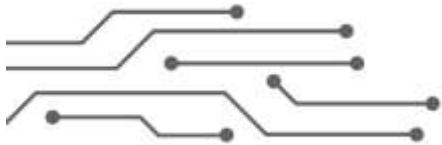
- Sprawdź, które z Twoich dokumentów, które są często używane w procesach, mogą zostać zdigitalizowane przy użyciu bibliotek OCR i przekonwertowane do postaci tabelarycznej.
- Sprawdź narzędzia do inżynierii odwrotnej, aby wyodrębnić dane z CAD (BIM).
- Spróbuj skonfigurować automatyczne pobieranie danych z dokumentów lub formatów często używanych w pracy do formularza tabeli
- Konfiguracja automatycznej konwersji pomiędzy różnymi formatami danych

■ Tworzenie bazy wiedzy do kategoryzacji

- Opracowanie wewnętrznego lub wykorzystanie istniejącego klasyfikatora elementów zgodnego ze standardami branżowymi.
- Dokumentowanie wzajemnych powiązań między różnymi systemami klasyfikacji
- Przedyskutuj ze swoim zespołem temat stosowania ujednoliconego systemu identyfikacji i jednoznacznej klasyfikacji elementów.
- Rozpocznij tworzenie procesu automatycznej walidacji danych - zarówno tych, z którymi pracujesz w zespole, jak i tych, które są przekazywane do systemów zewnętrznych.

Korzystając z tych podejść, można znacznie poprawić jakość danych i uprościć ich późniejsze przetwarzanie i przekształcanie. W kolejnych częściach tej książki przyjrzymy się, jak zastosować już ustrukturyzowane i przygotowane dane do zautomatyzowanych obliczeń, szacowania kosztów, tworzenia harmonogramów i zarządzania projektami budowlanymi.





IV CZĘŚĆ

KALKULACJE KOSZTÓW I CZASU: WŁĄCZANIE DANYCH DO PROCESÓW BUDOWLANYCH

Część piąta poświęcona jest praktycznym aspektom wykorzystania danych do optymalizacji kosztorysowania i planowania przedsięwzięć budowlanych. Szczegółowo przeanalizowano metodę szacowania kosztów opartą na zasobach oraz automatyzację procesów szacowania. Rozważane są metody automatycznego pozyskiwania Quantity Take-Off) z modeli CAD (BIM) oraz ich integracja z systemami obliczeniowymi. Zbadano technologie modelowania 4D i 5D dla harmonogramowania czasu i zarządzania kosztami budowy, wraz z konkretnymi przykładami ich zastosowania. Przedstawiono analizę rozszerzonych warstw informacyjnych 6D -8D, zapewniających zintegrowane podejście do oceny zrównoważonego rozwoju, eksploatacji i bezpieczeństwa obiektów nieruchomości. Szczegółowo omówiono metody obliczania śladu węglowego i wskaźników ESG projektów budowlanych w kontekście współczesnych wymagań i standardów środowiskowych. Krytycznie oceniono możliwości i ograniczenia tradycyjnych systemów ERP i PMIS w zarządzaniu procesami budowlanymi, wraz z analizą ich wpływu na przejrzystość cen. Prognozowane są perspektywy przejścia od rozwiązań zamkniętych do otwartych standardów i elastycznych narzędzi analizy danych, zdolnych do zapewnienia większej efektywności procesów budowlanych.

ROZDZIAŁ 4.1.

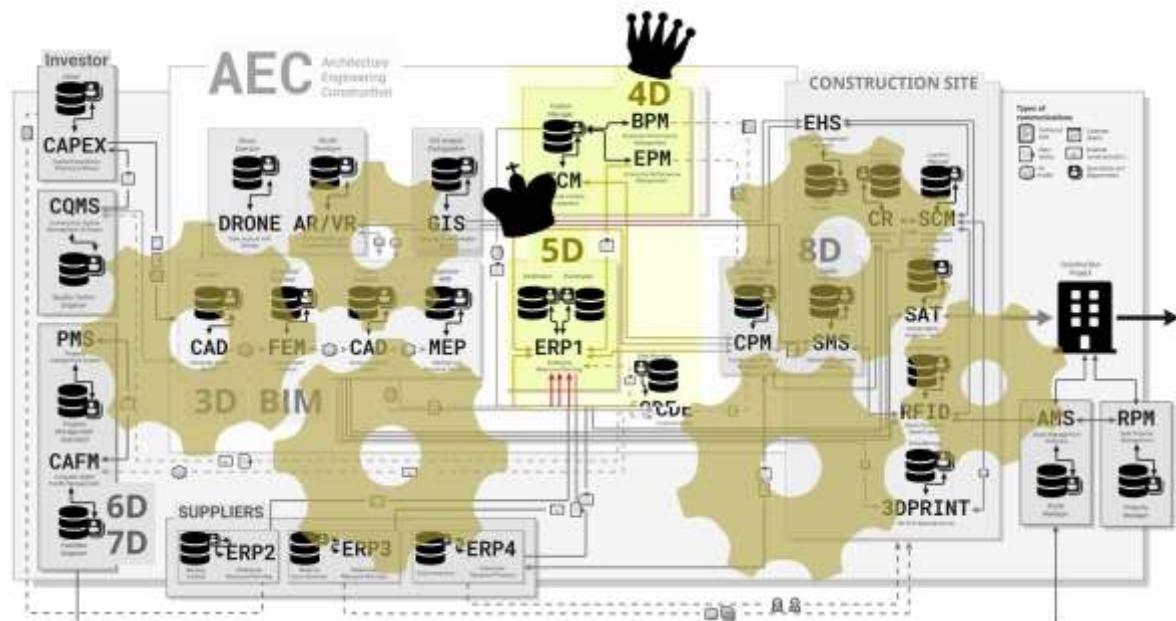
KALKULACJE KOSZTÓW I SZACUNKI DLA PROJEKTÓW BUDOWLANYCH

Podstawy budownictwa: szacowanie ilości, kosztów i czasu

Wśród wielu procesów biznesowych, które decydują o trwałości firmy w branży budowlanej, szczególne znaczenie - podobnie jak tysiące lat temu - mają procesy dokładnego szacowania liczby elementów, kosztów projektu i czasu realizacji (rys. 5.1-1).

Rozwój pisma był wynikiem wielu czynników, w tym potrzeby rejestrowania transakcji gospodarczych, handlu i zarządzania zasobami we wczesnych społeczeństwach. Pierwsze dokumenty o znaczeniu prawnym, gliniane tabliczki z obliczeniami kosztów materiałów i robocizny, były używane w kontekście handlu i budownictwa. Tabliczki te rejestrowały zobowiązania stron w zakresie budowy struktur i były przechowywane jako dowód umów oraz relacji pieniężnych i handlowych.

Przez tysiące lat podejście do szacowania pozostawało w dużej mierze niezmienione: obliczenia były wykonywane ręcznie, opierając się na doświadczeniu i intuicji inżyniera szacującego. Jednak wraz z pojawieniem się modułowych systemów ERP i narzędzi CAD, tradycyjne podejście do szacowania ilości, kosztów i czasu zaczęło się szybko zmieniać. Dzisiejsze technologie cyfrowe pozwalają w pełni zautomatyzować kluczowe obliczenia czasu i kosztów, umożliwiając większą dokładność, szybkość i przejrzystość w planowaniu zasobów projektów budowlanych.



Rys. 5.1-1 Spośród wielu różnych systemów, narzędzia odpowiedzialne za pomiary ilości, kosztów i czasu są najważniejsze w biznesie.

Firmy budowlane koncentrują się na dokładnych danych dotyczących czasu i kosztów. Te z kolei zależą od ilości wykorzystanych materiałów i robocizny, a ich przejrzystość wpływa na rentowność. Jednak złożoność procesów obliczeniowych i ich brak przejrzystości często prowadzą do wyższych kosztów projektu, niedotrzymania terminów, a nawet bankructwa.

Według raportu KPMG "Familiar Problems - New Approaches" (2023), tylko 50% projektów budowlanych jest kończonych na czas, a 87% firm zgłasza zwiększoną kontrolę nad ekonomicą projektów kapitałowych. Główne problemy związane są z brakiem wykwalifikowanego personelu i trudnościami w prognozowaniu ryzyka [2].

Historyczne dane dotyczące kosztów i czasu trwania procesów są gromadzone podczas budowy poprzednich projektów przez cały okres istnienia firmy budowlanej i wprowadzane do baz danych różnych systemów (ERP, PMIS BPM, EPM itp.).

Posiadanie dobrej jakości historycznych danych kosztowych jest główną przewagą konkurencyjną dla organizacji budowlanej, bezpośrednio wpływając na jej przetrwanie.

Działy szacowania i kosztorysowania w firmach budowlanych i inżynierijnych są tworzone w celu gromadzenia, przechowywania i aktualizowania danych historycznych dotyczących kalkulacji projektów. Ich główną funkcją jest gromadzenie i systematyzowanie doświadczenia firmy, co pozwala z czasem poprawić dokładność szacowania zakresu, harmonogramu i kosztów nowych projektów. Takie podejście pomaga zminimalizować błędy w przyszłych obliczeniach opartych na praktyce i wynikach już zrealizowanych projektów.

Metody obliczania szacunkowych kosztów projektów

Specjalisi ds. kosztorysowania stosują różne metody szacowania, z których każda koncentruje się na określonym rodzaju danych, dostępności informacji i poziomie szczegółowości projektu. Najpopularniejsze z nich obejmują:

- **Metoda oparta na zasobach:** szacowanie szacunkowego kosztu projektu w oparciu o szczegółową analizę wszystkich wymaganych zasobów, takich jak materiały, sprzęt i robocizna. Metoda ta wymaga szczegółowej listy wszystkich zadań i zasobów wymaganych do wykonania każdego zadania, a następnie obliczenia ich kosztu. Metoda ta jest bardzo dokładna i szeroko stosowana w szacowaniu kosztów.
- **Metoda parametryczna:** wykorzystuje modele statystyczne do szacowania kosztów na podstawie parametrów projektu. Może to obejmować analizę kosztu na jednostkę miary, taką jak powierzchnia budynku lub zakres prac, i dostosowanie tych wartości do konkretnych warunków projektu. Metoda ta jest szczególnie skuteczna na wczesnych etapach, gdy szczegółowe informacje nie są jeszcze dostępne.
- **Metoda jednostkowa (metoda kosztu jednostkowego):** oblicza szacunkowy koszt projektu w oparciu o koszt na jednostkę miary (np. na metr kwadratowy lub metr sześcienny). Zapewnia to szybki i łatwy sposób porównywania i analizowania kosztów różnych projektów lub ich części.
- **Ocena ekspercka (metoda Delphi):** oparta na opiniach ekspertów, którzy wykorzystują swoje doświadczenie i wiedzę do oszacowania wartości projektu. Podejście to jest przydatne, gdy dokładne dane bazowe nie są dostępne lub projekt jest wyjątkowy.

Warto zauważyć, że metoda parametryczna i oceny ekspertów mogą być dostosowane do modeli uczenia maszynowego. Pozwala to na automatyczne generowanie prognoz kosztów i harmonogramów projektu na podstawie próbek szkoleniowych. Przykłady zastosowania takich modeli omówiono bardziej szczegółowo w rozdziale "Przykład wykorzystania uczenia maszynowego do znajdowania kosztów i harmonogramów projektu" (rys. 9.3-5).

Niemniej jednak, metoda oparta na zasobach pozostaje najbardziej popularną i szeroko stosowaną w praktyce światowej. Zapewnia ona nie tylko dokładną ocenę szacowanego kosztu, ale także pozwala obliczyć czas trwania poszczególnych procesów na placu budowy i całego projektu jako całości (więcej szczegółów w rozdziale "Harmonogramy budowy i 4D - dane projektowe").

Metoda oparta na zasobach kosztorysowanie i szacowanie w budownictwie

Rachunek kosztów oparty na zasobach to metoda rachunkowości zarządczej, w której koszt projektu opiera się na bezpośrednim rozliczeniu wszystkich zaangażowanych zasobów. W budownictwie podejście to obejmuje szczegółową analizę i ocenę wszystkich materiałów, robocizny i zasobów technicznych wymaganych do wykonania prac.

Metoda oparta na zasobach, zapewnia wysoki stopień przejrzystości i dokładności w planowaniu budżetu, ponieważ koncentruje się na rzeczywistych cenach zasobów w momencie szacowania. Jest to szczególnie ważne w niestabilnym środowisku gospodarczym, w którym wahania cen mogą znacząco wpływać na całkowity koszt projektu.

W kolejnych rozdziałach przyjrzymy się szczegółowo procesowi kalkulacji kosztów w oparciu o zasoby. Aby lepiej zrozumieć jego zasady w budownictwie, narysujemy analogię do kalkulacji kosztów kolacji w restauracji. Menedżer restauracji, planując wieczór, sporządza listę niezbędnych produktów, bierze pod uwagę czas gotowania każdego dania, a następnie mnoży koszty przez liczbę gości. W budownictwie proces jest podobny: dla każdej kategorii elementów projektu (obiektów) generowane są szczegółowe szacunki Recipes, a całkowity koszt projektu jest określany poprzez zsumowanie wszystkich kosztów w całkowitym rachunku - ostatecznym szacunku według kategorii.

Kluczowym i początkowym etapem podejścia opartego na zasobach jest stworzenie początkowej bazy danych firmy. Na pierwszym etapie kalkulacji kosztów tworzona jest uporządkowana lista wszystkich przedmiotów, materiałów, rodzajów pracy i zasobów, którymi firma dysponuje w ramach swoich projektów budowlanych - od gwoździa w magazynie po opis ludzi poprzez ich kwalifikacje i stawkę godzinową. Informacje te są usystematyzowane w jednej "Bazie Danych Zasobów i Materiałów Budowlanych" - rejestrze tabelarycznym zawierającym dane dotyczące nazw, cech, jednostek miary i aktualnych cen. To właśnie ta baza danych staje się głównym i podstawowym źródłem informacji dla wszystkich późniejszych obliczeń zasobów - zarówno kosztów, jak i harmonogramu prac.

Baza danych zasobów budowlanych: katalog materiałów i robót budowlanych

Baza danych lub tabela zasobów i materiałów budowlanych - zawiera szczegółowe informacje o każdym elemencie, który może być wykorzystany w projekcie budowlanym - produkcie, przedmiocie, materiale lub usłudze, w tym jego nazwę, opis, jednostkę miary i koszt jednostkowy, zapisane w ustrukturyzowanej

formie. W tabeli tej można znaleźć wszystko, od różnych rodzajów paliw i materiałów wykorzystywanych w projektach po szczegółowe listy specjalistów w postaci różnych kategorii z opisami stawek godzinowych (rys. 5.1-2).

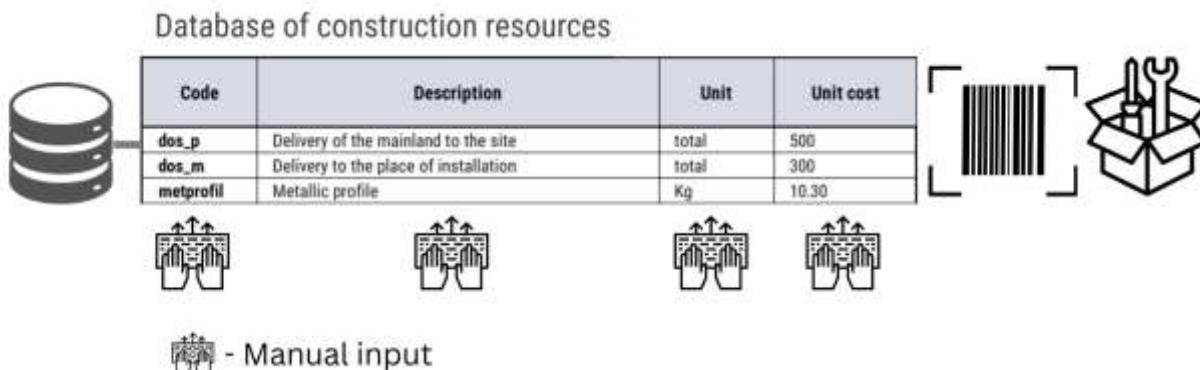
Database of resources		
	1st grade potatoes 1 kg \$2,99	
	Black Angus marble beef 1 kg \$26,99	
	Broccoli 1 pcs \$1,99	
		
	Sand lime bricks 1 pcs \$1	
		
	JCB 3CX backhoe loader 1 h \$150	
		
	Laborer of the 1st category 1 h \$30	

Rys. 5.1-2 Tabela zasobów to lista składników opisująca materiał i usługę wraz z kosztem jednostkowym.

"Baza zasobów" przypomina katalog produktów w sklepie internetowym, w którym każda pozycja posiada szczegółowy opis swoich atrybutów. Ułatwia to kosztorysantom wybór odpowiednich zasobów (jak wybór produktów podczas dodawania do koszyka) potrzebnych do obliczenia konkretnych procesów budowlanych w formie obliczeń (ostateczne zamówienie w sklepie internetowym).

Bazę danych zasobów można również traktować jako listę wszystkich składników w książce kucharskiej restauracji. Każdy materiał budowlany, sprzęt i usługa są podobne do składników używanych w przepisach. "Baza danych zasobów" to szczegółowa lista wszystkich składników - materiałów budowlanych i usług, w tym ich koszt jednostkowy: sztuka, metr, godzina, litr itp.

Nowe elementy encji mogą być dodawane do tabeli "Bazy danych zasobów budowlanych" na dwa sposoby - ręcznie (rys. 5.1-3) lub automatycznie poprzez integrację z systemami zarządzania zapasami firmy lub bazami danych dostawców.



Rys. 5.1-3 Baza danych zasobów jest wypełniana ręcznie lub automatycznie pobiera dane z innych baz danych.

Typowa firma budowlana średniej wielkości korzysta z bazy danych zawierającej tysiące, a czasem dziesiątki tysięcy pozycji ze szczegółowymi opisami, które można wykorzystać w projektach budowlanych. Dane te są następnie automatycznie wykorzystywane w umowach i dokumentacji projektowej w celu dokładnego opisania zakresu prac i procesów

Aby nadążyć za zmieniającymi się warunkami rynkowymi, takimi jak inflacja, atrybut "koszt jednostkowy" dla każdego produktu (towaru lub usługi) w bazie danych zasobów (Rys. 5.1-3) jest regularnie aktualizowany ręcznie lub poprzez automatyczne pobieranie aktualnych cen z innych systemów lub platform internetowych.

Aktualizacja kosztu jednostkowego zasobu może odbywać się co miesiąc, co kwartał lub co rok - w zależności od charakteru zasobu, inflacji i zewnętrznego klimatu gospodarczego. Takie aktualizacje są niezbędne do utrzymania dokładności obliczeń i szacunków, ponieważ te podstawowe elementy są punktem wyjścia dla pracy kosztorysantów. Aktualne dane są wykorzystywane do tworzenia szacunków, budżetów i harmonogramów, które odzwierciedlają rzeczywiste warunki rynkowe i zmniejszają ryzyko błędów w późniejszych obliczeniach projektu.

Kompilacja obliczeń i kosztorysowanie prac w oparciu o bazę zasobów

Po wypełnieniu "Bazy danych zasobów budowlanych" (rys. 5.1-3) minimalnymi jednostkami można rozpocząć tworzenie obliczeń, które są obliczane dla każdego procesu lub pracy na placu budowy dla określonych jednostek miary: na przykład dla jednego metra sześciennego betonu, jednego metra kwadratowego ściany z płyt gipsowo-kartonowych, metra krawężnika lub instalacji okna.

Na przykład, aby zbudować ścianę z cegieł o powierzchni 1 m² (Rys. 5.1-4), w oparciu o doświadczenie z poprzednich projektów, wymagane jest około 65 cegieł (jednostka "Cegła silikatowa") w cenie 1 USD za sztukę (atribut "Koszt za sztukę"), co daje łącznie 65 USD. Ponadto, zgodnie z moim doświadczeniem, wymagane jest użycie sprzętu budowlanego (jednostka "JCB 3CX Loader") przez 10 minut, który umieści cegły w pobliżu obszaru roboczego. Ponieważ wynajęcie sprzętu kosztuje 150 USD za godzinę, 6 minut użytkowania kosztowałoby około 15 USD. Ponadto, wykonawca układania cegieł będzie potrzebny przez 2 godziny, ze stawką godzinową w wysokości 30 USD i łączną kwotą 60 USD.

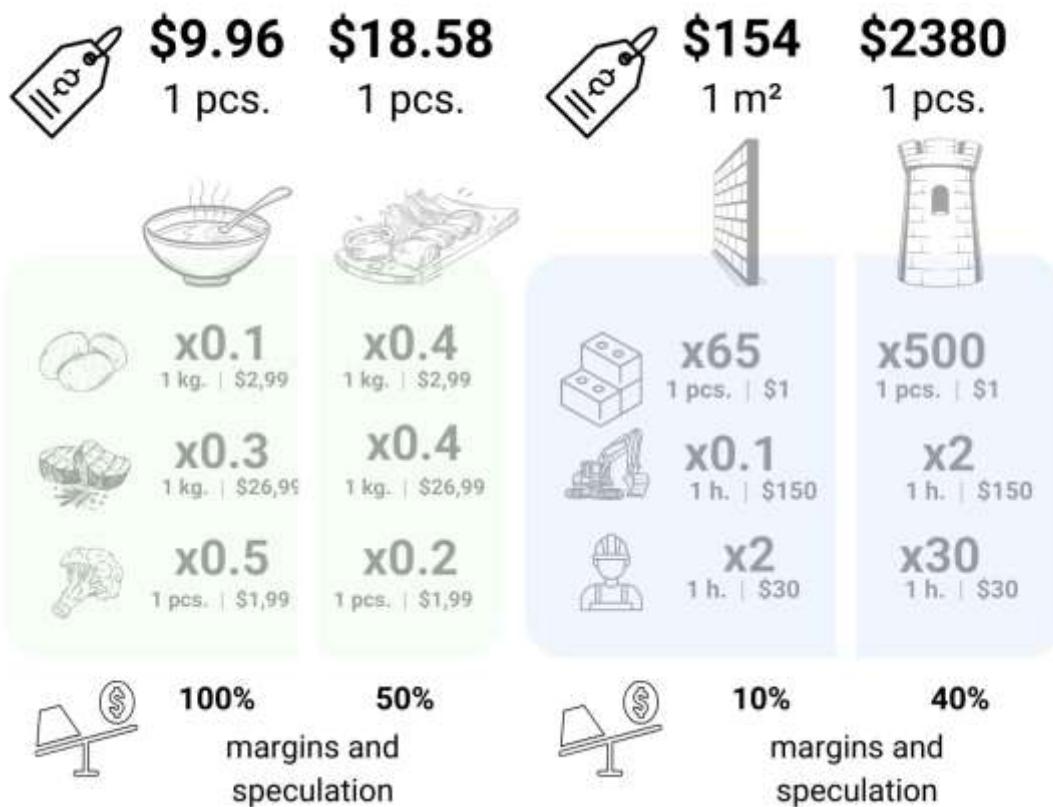
x0.1 1 kg \$2,99	x0.4 1 kg. \$2,99	x65 1 pcs \$1	x500 1 pcs \$1
x0.3 1 kg \$26,99	x0.4 1 kg. \$26,99	x0.1 1 h \$150	x2 1 h \$150
x0.5 1 pcs \$1,99	x0.2 1 pcs. \$1,99	x2 1 h \$30	x30 1 h \$30
+ margins and speculation	+ margins and speculation	+ margins and speculation	+ margins and speculation

Rys. 5.1-4 Kalkulacje kosztów zawierają szczegółową listę materiałów budowlanych i usług wymaganych do wykonania robót i procesów.

Skład obliczeń (tzw. "receptur") jest tworzony na podstawie historycznych doświadczeń zgromadzonych przez firmę w procesie wykonywania dużej ilości podobnych prac. To praktyczne doświadczenie jest zwykle gromadzone poprzez informacje zwrotne z placu budowy. W szczególności brygadzista zbiera informacje bezpośrednio na placu budowy, rejestrując rzeczywiste koszty pracy, zużycie materiałów i niuanse operacji technologicznych. We współpracy z działem kosztorysowania, informacje te są następnie iteracyjnie udoskonalane: opisy procesów są udoskonalane, zestaw zasobów jest dostosowywany, a kosztorysy są aktualizowane w celu odzwierciedlenia rzeczywistych danych z ostatnich projektów.

Tak jak przepis opisuje składniki i ilości potrzebne do przygotowania potrawy, tak arkusz kalkulacyjny zawiera szczegółową listę wszystkich materiałów budowlanych, zasobów i usług wymaganych do wykonania określonego zadania lub procesu.

Regularnie wykonywana praca pozwala pracownikom, brygadistom i kosztorysantom zorientować się w wymaganej ilości zasobów: materiałów, paliwa, czasu pracy i innych parametrów niezbędnych do wykonania jednostki pracy (rys. 5.1-5). Dane te są wprowadzane do systemów szacowania w formie tabel, w których każde zadanie i operacja są opisane za pomocą minimalnych elementów bazy zasobów (ze stale aktualizowanymi cenami), co zapewnia dokładność obliczeń.



Rys. 5.1-5 Stawki jednostkowe są zbierane dla każdego zadania, gdzie atrybut objętości podmiotu jest mnożony przez jego ilość z dodanym procentem zysku.

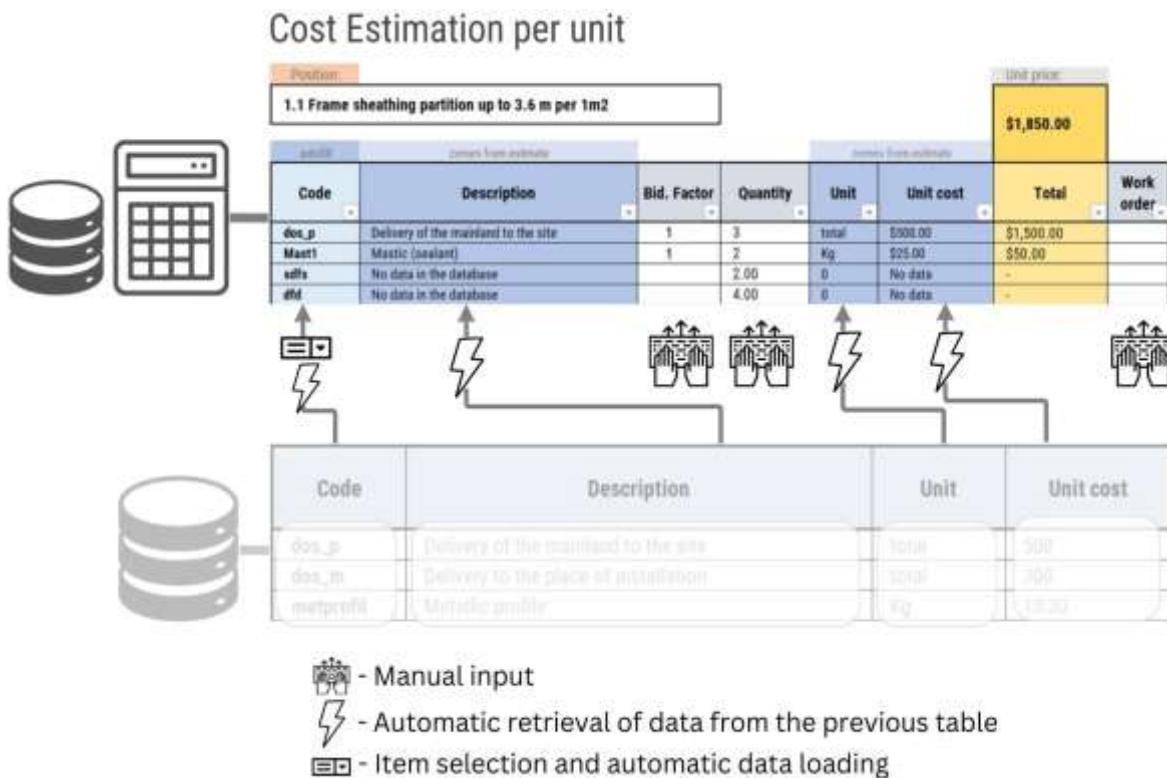
Aby uzyskać całkowity koszt każdego procesu lub działania (obiektu kalkulacji kosztów), atrybut kosztu jest mnożony przez jego liczbę i współczynniki. Współczynniki mogą uwzględniać różne czynniki, takie jak złożoność pracy, charakterystyka regionalna, stopa inflacji, potencjalne ryzyko (oczekiwany procent kosztów ogólnych) lub spekulacje (dodatkowy współczynnik zysku).

Estymator, jako analityk, przekształca doświadczenie i zalecenia brygadzisty w znormalizowane szacunki, opisujące procesy budowlane poprzez jednostki zasobów w formie tabelarycznej. Zasadniczo zadaniem estymatora jest gromadzenie i strukturyzowanie, za pomocą parametrów i współczynników, informacji pochodzących z placu budowy.

W związku z tym ostateczny koszt za jednostkę pracy (np. metr kwadratowy lub sześcienny lub jedną instalację pojedynczego urządzenia) obejmuje nie tylko bezpośrednie koszty materiałów i robocizny, ale także marże firmy, koszty ogólne, ubezpieczenie i inne czynniki (Rys. 5.1-6)

Jednocześnie nie musimy już martwić się o rzeczywiste ceny w obliczeniach (receptur), ponieważ rzeczywiste ceny są zawsze odzwierciedlane w "bazie zasobów" (tabeli składników). Na poziomie obliczeń dane z bazy zasobów są automatycznie ładowane do tabeli (np. według kodu artykułu lub jego unikalnego identyfikatora), która ładuje opis i rzeczywistą cenę za jednostkę, która z kolei może być automatycznie ładowana z platform internetowych lub sklepu internetowego z materiałami budowlanymi. Estymator na

poziomie obliczeń musi jedynie opisać pracę lub proces za pomocą atrybutu "ilość zasobów" i dodatkowych czynników.



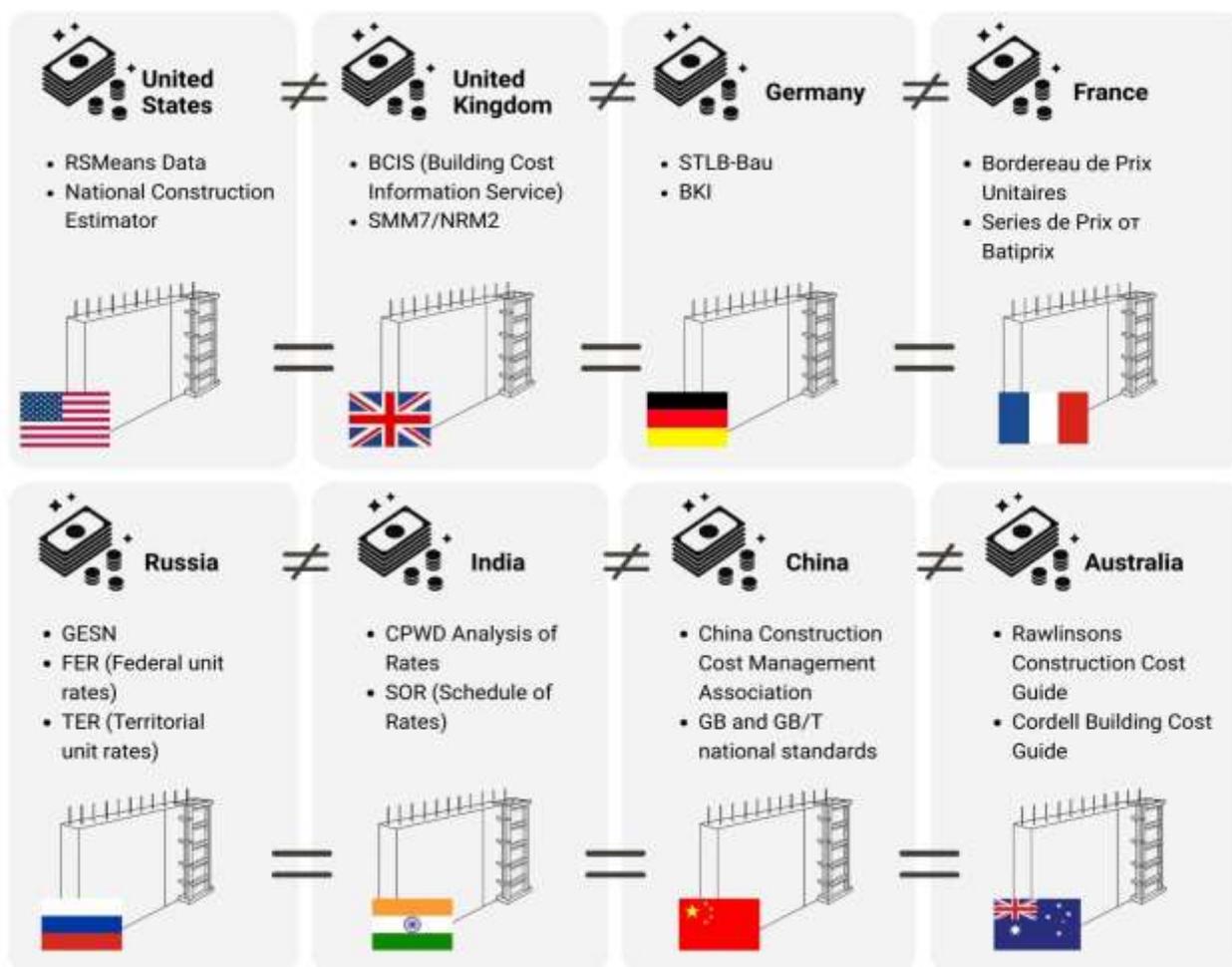
Rys. 5.1-6 Na etapie obliczania jednostkowego kosztu pracy wypełniane są tylko atrybuty liczby wymaganych zasobów, wszystko inne jest automatycznie ładowane z bazy danych zasobów.

Utworzone kosztorysy zadań są przechowywane w formie tabel szablonów typowych projektów, które są bezpośrednio powiązane z bazą danych zasobów budowlanych i materiałów. Szablony te stanowią znormalizowane receptury dla powtarzalnych typów pracy dla przyszłych projektów, zapewniając jednolitość obliczeń w całej firmie.

Gdy koszt dowolnego zasobu zmienia się w bazie danych (rys. 5.1-3) - ręcznie lub automatycznie poprzez pobieranie aktualnych cen rynkowych (np. w warunkach inflacji) - aktualizacje są natychmiast odzwierciedlane we wszystkich powiązanych kosztorysach (rys. 5.1-6). Oznacza to, że tylko baza zasobów musi zostać zmieniona, podczas gdy szablony kalkulacji kosztów i szacunki pozostają niezmienione w czasie. Takie podejście zapewnia stabilność i powtarzalność obliczeń dla wszelkich wahań cen, które są uwzględniane tylko w stosunkowo prostej tabeli zasobów (Rys. 5.1-3).

Dla każdego nowego projektu tworzona jest kopia standardowego szablonu kalkulacji kosztów, co pozwala na wprowadzanie zmian i dostosowywanie działań do określonych wymagań bez zmiany oryginalnego szablonu przyjętego przez firmę. Takie podejście zapewnia elastyczność w dostosowywaniu obliczeń: można wziąć pod uwagę specyfikę placu budowy, życzenia klienta, wprowadzić współczynniki ryzyka lub rentowności (spekulacji) - wszystko bez naruszania standardów firmy. Pomaga to firmie znaleźć równowagę między maksymalizacją zysku, satysfakcją klienta i utrzymaniem konkurencyjności.

W niektórych krajach takie szablony kalkulacji kosztów, gromadzone przez dziesięciolecia, są standaryzowane na poziomie krajowym i stają się częścią krajowych standardów systemu kalkulacji kosztów budowy (Rys. 5.1-7).



Rys. 5.1-7 Różne kraje na całym świecie mają własne zasady kalkulacji kosztów z własnymi (normatywnymi) kompendiami i standardami dla prac budowlanych w zakresie kalkulacji kosztów tego samego elementu.

Takie znormalizowane bazy szacunków zasobów (rys. 5.1-7) są obowiązkowe do wykorzystania

wszystkich uczestników rynku, zwłaszcza w przypadku projektów finansowanych ze środków publicznych. Taka standaryzacja zapewnia przejrzystość, porównywalność i uczciwość w kształtowaniu cen i zobowiązań umownych dla klienta.

Ostateczny kosztorys projektu: od szacunków do budżetów

Państwowe i branżowe standardy szacowania odgrywają różne role w praktyce budowlanej w różnych krajach. Podczas gdy niektóre kraje wymagają ścisłego przestrzegania jednego standardu, większość rozwiniętych gospodarek przyjmuje bardziej elastyczne podejście. W gospodarkach rynkowych, rządowe standardy budowlane zazwyczaj służą jedynie jako punkt odniesienia. Firmy budowlane dostosowują te standardy do swoich modeli operacyjnych lub całkowicie je zmieniają, uzupełniając je o własne, niestandardowe czynniki. Korekty te odzwierciedlają doświadczenie korporacyjne, efektywność zarządzania zasobami i często czynniki, w których na przykład można uwzględnić zyski spekulacyjne firmy.

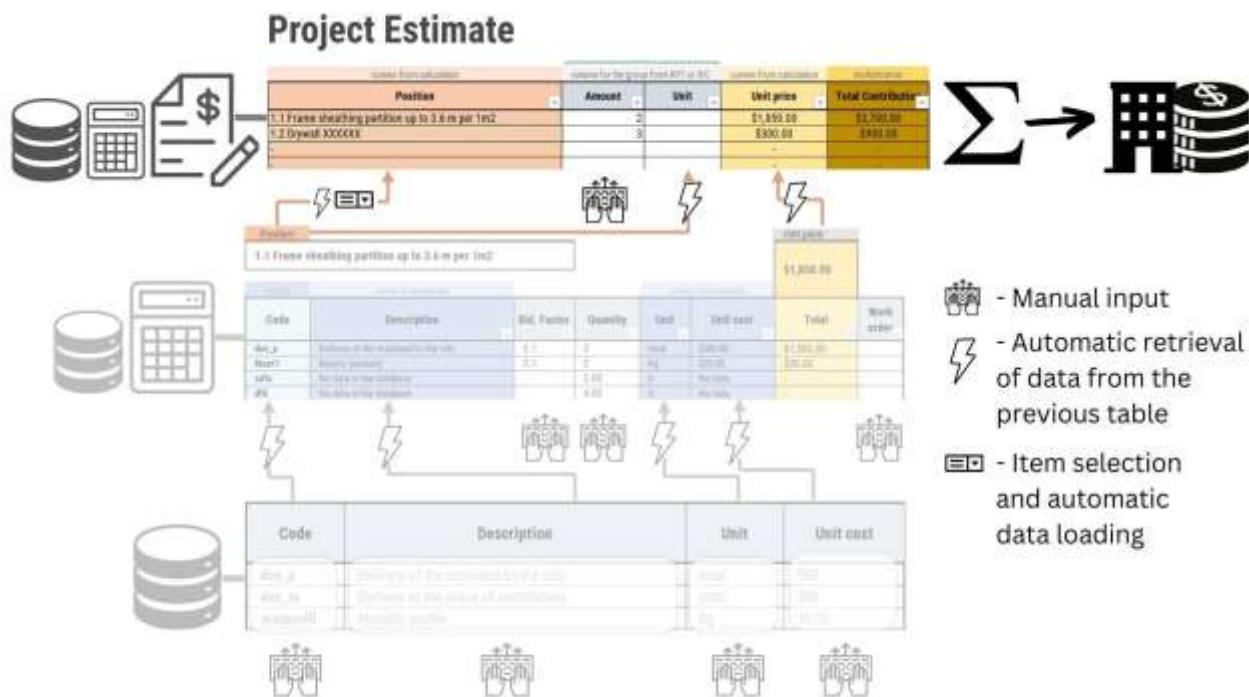
W rezultacie poziom konkurencji, popyt rynkowy, docelowe marże, a nawet relacje z konkretnymi klientami mogą prowadzić do znacznych odchyleń od standardowych norm. Praktyka ta zapewnia elastyczność rynku, ale także utrudnia przejrzyste porównanie ofert od różnych wykonawców, wprowadzając element spekulacyjnej wyceny do branży budowlanej na tym etapie procesu obliczeniowego.

Po przygotowaniu szablonów kalkulacji dla poszczególnych działań i procesów - lub, częściej, po prostu skopiowaniu ze standardowych szacunków rządowych (rys. 5.1-7) ze współczynnikami dodanymi w celu odzwierciedlenia "specyfiki" konkretnej firmy - ostatnim krokiem jest pomnożenie kosztu każdej pozycji przez odpowiedni atrybut zakresu prac lub procesów w nowym projekcie.

Przy obliczaniu całkowitego kosztu nowego projektu budowlanego, kluczowym krokiem jest podsumowanie kosztów wszystkich pozycji kosztorysowych, pomnożonych przez ilość tych elementów pracy w projekcie.

Aby utworzyć całkowity koszt projektu, w naszym uproszczonym przykładzie zaczynamy od obliczenia kosztu budowy jednego metra kwadratowego ściany i mnożymy koszt jego obliczenia (np. praca "1m² standardowa instalacja elementów sciennych") przez całkowitą liczbę metrów kwadratowych ścian w projekcie (np. atrybut "Powierzchnia" lub "Ilość" (rys. 5.1-8) encji typu "Elementy scienne" z CAD projektu lub obliczenia brygadzisty).

Podobnie obliczamy koszt dla wszystkich elementów projektu (rys. 5.1-8): bierzemy koszt na jednostkę pracy i mnożymy go przez objętość danego elementu lub grupy elementów w projekcie. Estymator musi jedynie wprowadzić liczbę tych elementów, działań lub procesów w projekcie na stronie w formie objętości lub ilości. Pozwala to na automatyczne wygenerowanie kompletnego kosztorysu budowlanego.



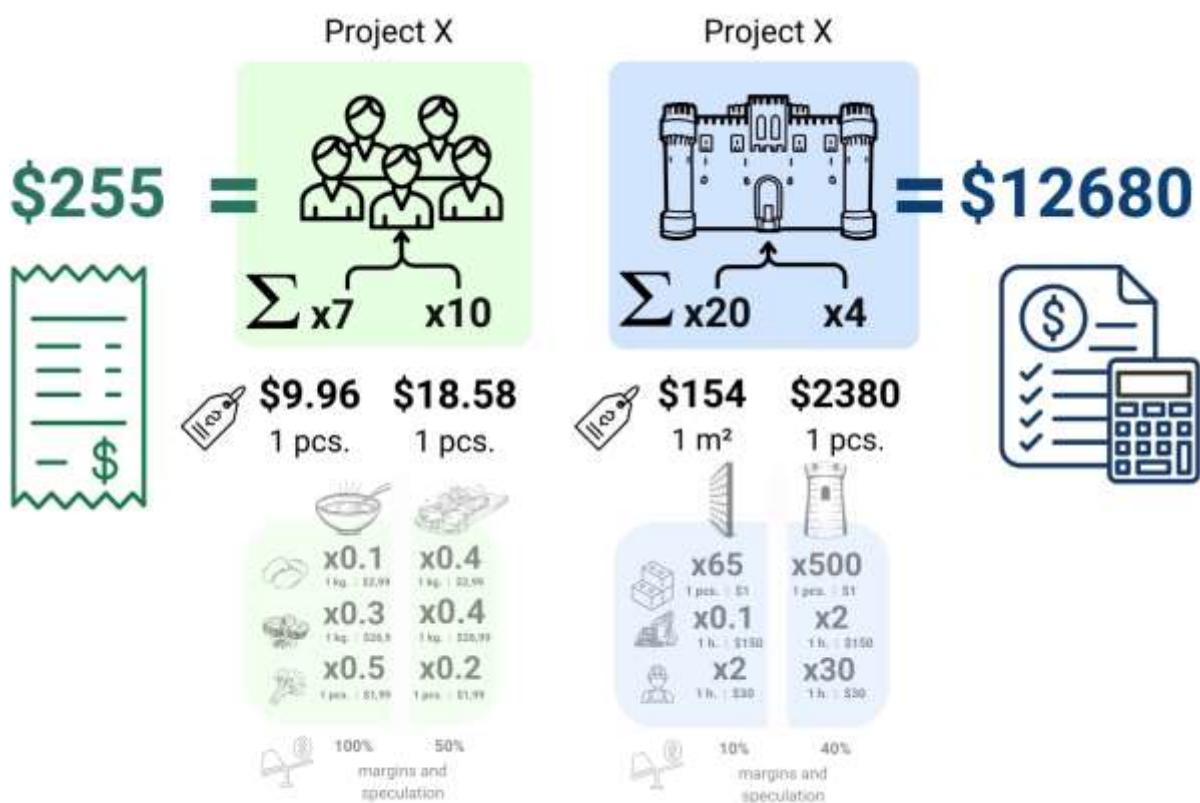
Rys. 5.1-8 Na etapie tworzenia kosztorysu wprowadzamy jedynie zakres prac.

Podobnie jak w przypadku kalkulacji, na tym poziomie wgrywamy automatycznie gotowe pozycje kalkulacyjne (z szablonu kalkulacji lub nowe skopiowane z szablonu i edytowane), które automatycznie niosą ze sobą aktualny koszt jednostki pracy (który jest aktualizowany automatycznie z bazy zasobów (Rys. 5.1-8 dolna tabela)). W związku z tym, w przypadku jakiekolwiek zmiany danych w bazie danych zasobów lub tabelach kosztorysowych - dane w kosztorysie zostaną automatycznie zaktualizowane na bieżący dzień, bez konieczności zmiany kosztorysu lub samego kosztorysu.

W kontekście restauracji, ostateczny koszt wydarzenia jest obliczany w podobny sposób i jest równy ostatecznemu kosztowi całej kolacji, gdzie koszt każdego dania pomnożony przez liczbę gości sumuje się do całkowitego kosztu czeku (Rys. 5.1-9). I podobnie jak w budownictwie, przepisy na gotowanie w restauracji mogą się zmieniać przez dziesięciolecia. W przeciwnieństwie do cen, gdzie koszt składników może zmieniać się co godzinę.

Tak jak właściciel restauracji mnoży koszt każdego posiłku przez liczbę porcji i osób, aby określić całkowity koszt wydarzenia, tak kierownik ds. szacowania kosztów sumuje koszty wszystkich elementów projektu, aby uzyskać kompletny kosztorys budowlany.

W ten sposób dla każdego działania w projekcie określany jest jego ostateczny koszt (rys. 5.1-9), który pomnożony przez wielkość atrybutu podmiotu odpowiadającego temu działaniu - daje koszt grup działań, z których uzyskuje się ostateczny koszt całego projektu.



Rys. 5.1-9 Ostateczne oszacowanie jest obliczane poprzez zsumowanie atrybutu kosztu pracy każdego elementu przez jego atrybut zakresu.

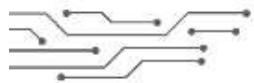
Całkowity koszt projektu (Rys. 5.1-8) zapewnia finansowy obraz projektu, umożliwiając klientom, inwestorom lub organizacjom finansującym zrozumienie całkowitego budżetu i zasobów finansowych wymaganych do realizacji projektu w danym dniu, biorąc pod uwagę aktualne ceny.

A jeśli procesy komplikowania baz zasobów, obliczeń i szacunków (receptury procesów) zostały już opracowane, półautomatyczne i dopracowane przez dziesiątki tysięcy lat i zarejestrowane na poziomie państwowym, to automatyczne uzyskiwanie informacji jakościowych o objętości i ilości elementów na ostatnim etapie ostatecznego oszacowania - dziś pozostaje wąskim gardłem w procesach wszystkich obliczeń kosztów i atrybutów czasowych projektu, a ogólnie ogólnego budżetu projektu.

Przez tysiące lat tradycyjną metodą obliczania objętości były ręczne metody pomiaru objętości i ilości przy użyciu płaskich rysunków. Wraz z nadaniem ery cyfrowej firmy odkryły, że informacje o objętości i ilości można teraz automatycznie wyodrębnić z danych geometrycznych zawartych w modelach CAD, rewolucjonizując tysiącletnie sposoby uzyskiwania danych ilościowych.

Nowoczesne podejście do szacowania i wyceny procesów obejmuje automatyczną ekstrakcję objętościowych i ilościowych atrybutów z baz danych CAD, które mogą być przesyłane i połączone z procesem kalkulacji kosztów w celu uzyskania aktualnych wielkości grupy projektowej na dowolnym

etapie projektowania aż do eksploatacji.



ROZDZIAŁ 4.2.

PRZEDMIAR I AUTOMATYCZNE TWORZENIE KOSZTORYSÓW I HARMONOGRAMÓW

Przejście z 3D do 4D i 5D: wykorzystanie parametrów objętościowych i ilościowych

Mając do dyspozycji tabele kosztorysowe z opisanymi procesami poprzez zasoby (rys. 5.1-8), kolejnym krokiem jest automatyczne uzyskanie parametrów objętościowych lub ilościowych dla grupy elementów, które są potrzebne do obliczeń i ostatecznego oszacowania.

Charakterystyki objętościowe elementów projektu - np. ścian lub płyt - mogą być automatycznie pobierane z baz danych CAD. Obiekty parametryczne utworzone w programach CAD są konwertowane za pomocą jądra geometrii na wartości liczbowe długości, szerokości, powierzchni, objętości i innych parametrów. Proces uzyskiwania objętości na podstawie geometrii 3D zostanie omówiony bardziej szczegółowo w kolejnej, szóstej części (rys. 6.3-3), poświęconej pracy z programami CAD (BIM). Oprócz objętości, liczbę podobnych elementów można również uzyskać z bazy danych modelu CAD poprzez filtrowanie i grupowanie obiektów według kategorii i właściwości. Te parametry, które umożliwiają grupowanie, stają się podstawą do łączenia elementów projektu poprzez obliczenia zasobów z obliczeniami, ostatecznymi szacunkami i budżetem całego projektu.

W ten sposób model danych wyodrębniony z modelu 3D (CAD) jest rozszerzany o nowe warstwy parametrów, oznaczane jako 4D i 5D. W nowych warstwach atrybutów encji, 4D (czas) i 5D (koszt), dane geometryczne 3D są wykorzystywane jako źródło wartości atrybutów objętości encji.

- **4D** to warstwa informacji o parametrach, która dodaje informacje o czasie trwania operacji budowlanych do parametrów 3D elementów. Dane te są niezbędne do planowania harmonogramów i zarządzania harmonogramami projektów
- **5D** to kolejny poziom rozszerzenia modelu danych, w którym elementy są uzupełniane o charakterystykę kosztową. Dodaje to aspekt finansowy do informacji geometrycznych: koszt materiałów, robót i sprzętu, co umożliwia obliczenia budżetowe, analizy rentowności i zarządzanie kosztami podczas procesu budowy.

Dane dotyczące kosztów i atrybutów 3D, 4D i 5D grup podmiotów projektu są opisane w sposób podobny do obliczeń w modułowych systemach ERP, PIMS (lub narzędziach podobnych do Excela) i są wykorzystywane do automatycznego kalkulowania kosztów i planowania budżetu zarówno poszczególnych grup, jak i całego budżetu projektu.

Atrybuty 5D i pobieranie objętości atrybutów z CAD

Podczas przygotowywania ostatecznego kosztorysu projektu budowlanego, którego przygotowanie omówiliśmy w poprzednich rozdziałach (Rys. 5.1-8), atrybuty objętości dla każdej kategorii elementów projektu są zbierane ręcznie lub pobierane ze specyfikacji atrybutów objętości dostarczanych przez oprogramowanie CAD.

Tradycyjna ręczna metoda obliczania ilości polega na tym, że brygadzista i kosztorysant analizują rysunki, które od tysięcy lat są przedstawiane jako linie na papierze, a od 30 lat w formatach cyfrowych, takich jak PDF (PLT) lub DWG. Opierając się na doświadczeniu zawodowym, mierzą ilości pracy i potrzebnych materiałów, często za pomocą linijki i kątomierza. Metoda ta wymaga znacznego wysiłku i czasu, a także szczególnej dbałości o szczegóły.

Określenie atrybutów zakresu w ten sposób może zajść od kilku dni do kilku miesięcy, w zależności od zakresu projektu. Ponadto, ponieważ wszystkie pomiary i obliczenia są wykonywane ręcznie, istnieje ryzyko błędu ludzkiego, który może prowadzić do niedokładnych danych, co z kolei wpływa na błędy w szacowaniu czasu i kosztów projektu, za które odpowiedzialność ponosi cała firma.

Nowoczesne metody oparte na wykorzystaniu baz danych CAD znacznie upraszczają obliczanie objętości. W modelach CAD geometria elementów zawiera już atrybuty objętości, które mogą być automatycznie obliczane (za pomocą jądra geometrii (rys. 6.3-3)) i prezentowane lub eksportowane w formie tabelarycznej.

W takim scenariuszu działa kosztorysowania prosi projektanta CAD o dane dotyczące ilości i objętości elementów projektu. Dane te są eksportowane jako arkusze kalkulacyjne lub bezpośrednio integrowane z bazami danych kosztorysowych - czy to Excel, ERP czy PMIS - systemy. Proces ten często rozpoczyna się nie od formalnego wniosku, ale od krótkiego dialogu między klientem (inicjatorem) a architektem kosztorysantem z firmy budowlanej lub projektowej. Poniżej znajduje się uproszczony przykład pokazujący, w jaki sposób ustrukturyzowana tabela do automatycznych obliczeń (QTO) jest tworzona na podstawie codziennej komunikacji:

- ⌚ Klient - "Chcę dodać kolejne piętro do budynku, w tej samej konfiguracji co pierwsze piętro".
- ⌚ Architekt (CAD) - "Dodanie trzeciego piętra, konfiguracja jest taka sama jak na drugim piętrze". Po tej wiadomości wysyła nową wersję CAD projektu do kosztorysanta.
- ⌚ Estymator automatycznie wykonuje grupowanie i obliczenia (ERP, PMIS, Excel) - "Przeprowadzę projekt przez arkusz kalkulacyjny Excel z regułami QTO (ERP, PMIS), uzyskam ilości według kategorii nowego piętra i wygeneruję kosztorys".

W rezultacie dialog tekstowy jest przekształcany w strukturę tabeli z regułami grupowania:

Element	Kategoria	Podłoga
Nakładanie się	OST_Floors	3

Element	Kategoria	Podłoga
Kolumna	OST_StructuralColumns	3
klatka schodowa	OST_Stairs	3

Po procesie automatycznego grupowania modelu CAD od projektanta zgodnie z zasadami QTO estymatora i automatycznego mnożenia objętości przez obliczenia zasobów (rys. 5.1-8) otrzymujemy następujące wyniki, które są wysyłane do klienta:

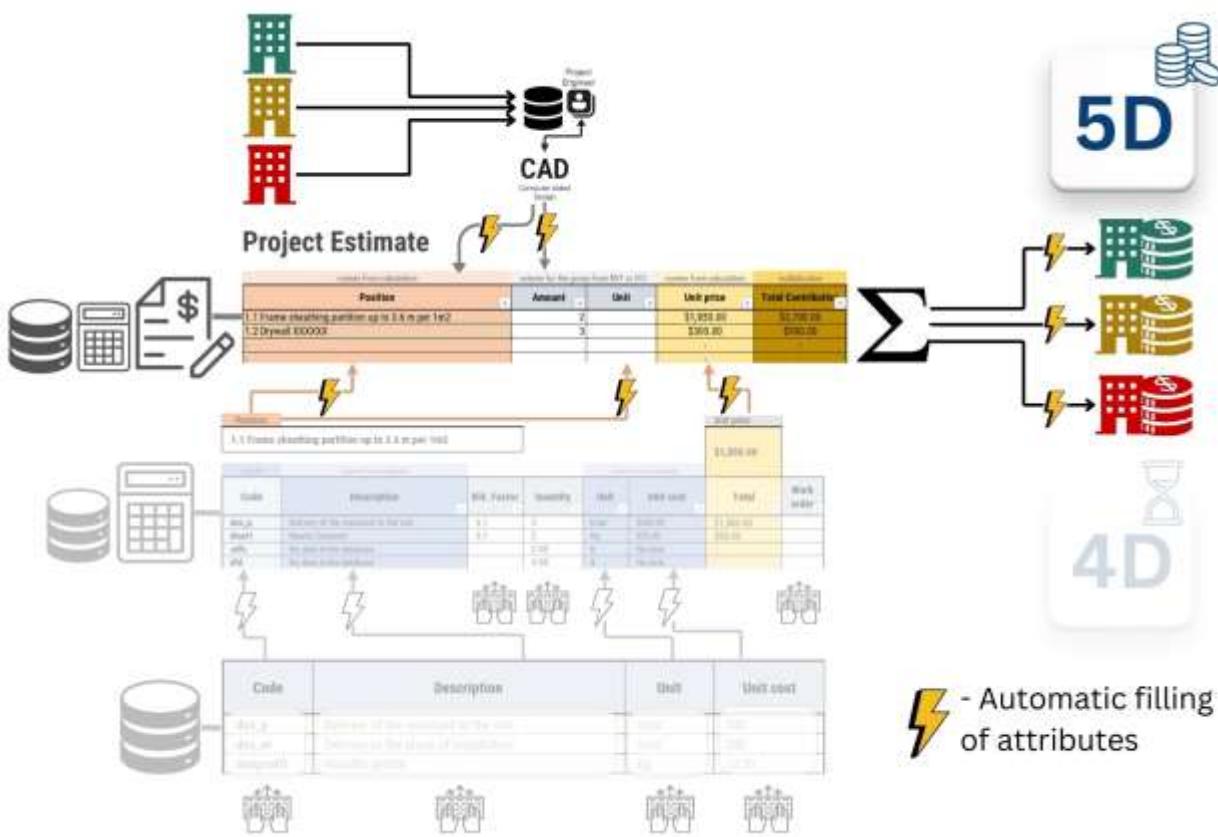
Element	Objętość	Podłoga	Cena za sztukę.	Całkowity koszt
Nakładanie się	420 M ²	3	150 €/M ²	63 000 €
Kolumna	4 szt.	3	2450 €/szt.	9 800 €
klatka schodowa	2 szt.	3	4.300 €/szt.	8 600 €
RAZEM:	-	-	-	81 400 €

💡 Klient - "Dziękuję, to dość dużo, musimy zmniejszyć kilka pokoi". I cykl powtarza się wiele razy.

Scenariusz ten może powtarzać się wielokrotnie, zwłaszcza w fazie zatwierdzania, gdzie klient oczekuje natychmiastowej informacji zwrotnej. W praktyce jednak takie procesy mogą ciągnąć się dniami, a nawet tygodniami. Obecnie, dzięki wprowadzeniu automatycznego grupowania i reguł obliczeniowych, czynności, które wcześniej zajmowały dużo czasu, powinny być wykonywane w ciągu kilku minut. Zautomatyzowane pozyskiwanie ilości, dzięki regułom grupowania, nie tylko przyspiesza obliczenia i szacunki, ale poprzez zminimalizowanie czynnika ludzkiego zmniejsza prawdopodobieństwo wystąpienia błędów, zapewniając przejrzystą i dokładną ocenę kosztów projektu.

Jeśli wymagania działu szacowania zostały początkowo uwzględnione podczas tworzenia modelu 3D w systemie CAD (co wciąż jest rzadkością w praktyce), a nazwy, identyfikatory grup elementów i ich atrybuty klasyfikacji są ustawione w postaci parametrów, które pokrywają się ze strukturami grup i klas szacowania, wówczas atrybuty objętościowe mogą być automatycznie przenoszone do systemów szacowania bez dodatkowych przekształceń.

Automatyczna ekstrakcja atrybutów objętościowych z CAD w postaci tabel specyfikacji pozwala szybko uzyskać aktualne dane na temat kosztów poszczególnych prac i projektu jako całości (rys. 5.2-1). Aktualizując tylko plik CAD z objętościami projektu w procesie obliczeniowym lub systemie obliczeniowym, firma może szybko przeliczyć kosztorys z uwzględnieniem najnowszych zmian, zapewniając wysoką dokładność i spójność wszystkich kolejnych obliczeń.



Rys. 5.2-1 Atrybuty objętości z tabel CAD lub baz danych są automatycznie wprowadzane do kosztorysu, umożliwiając natychmiastowe obliczenie całkowitego kosztu projektu.

Wraz z rosnącą złożonością projektów kapitałowych, obliczanie pełnego budżetu i analizowanie całkowitego kosztu projektów w takim scenariuszu (Rys. 5.2-1) - staje się kluczowym narzędziem do świadomego podejmowania decyzji.

Według badania Accenture Creating More Value through Capital Projects (2024) [20], wiodące firmy aktywnie integrują analitykę danych z inicjatywami cyfrowymi, wykorzystując informacje historyczne do przewidywania i optymalizacji wyników. Badania pokazują, że coraz więcej właścicieli-operatorów stosuje analitykę dużych zbiorów danych w celu przewidywania trendów rynkowych i oceny rentowności komercyjnej przed rozpoczęciem projektowania. Osiąga się to poprzez analizę hurtowni danych z istniejącego portfolio projektów. Ponadto, 79 procent właścicieli-operatorów wdraża "solidną" analitykę predykcyjną w celu oceny wydajności projektu i wspierania podejmowania decyzji operacyjnych w czasie rzeczywistym.

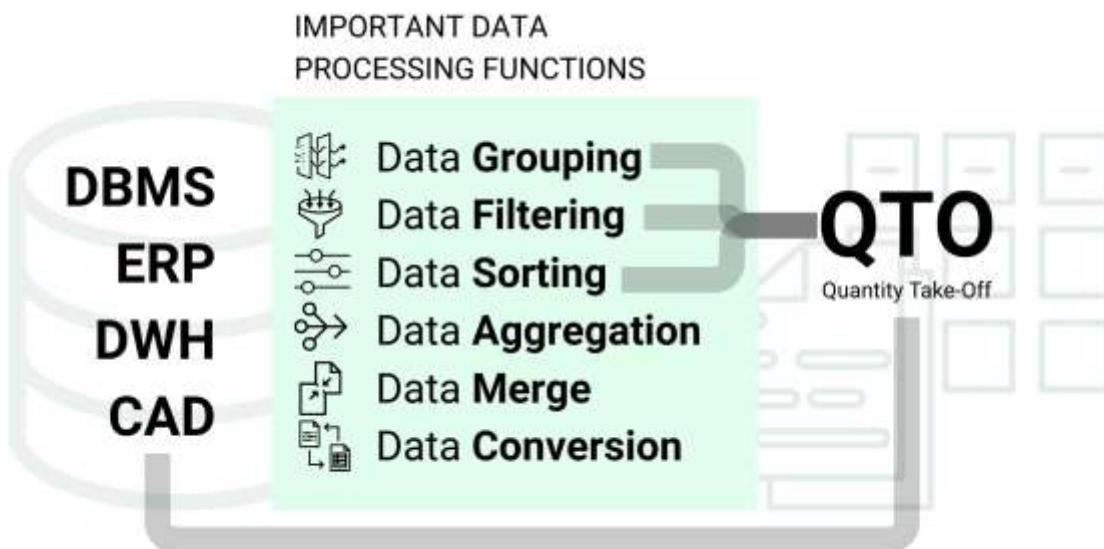
Nowoczesne efektywne zarządzanie projektami budowlanymi jest nierozerwalnie związane z przetwarzaniem i analizą dużych ilości informacji na wszystkich etapach projektowania i tych procesów, które poprzedzają projektowanie. Wykorzystanie hurtowni danych, obliczeń zasobów, modeli predykcyjnych i uczenia maszynowego pozwala nie tylko na minimalizację ryzyka w obliczeniach, ale także na podejmowanie strategicznych decyzji dotyczących finansowania projektu na wczesnych etapach projektowania. Więcej o hurtowniach danych i modelach predykcyjnych, które uzupełnią obliczenia, powiemy w dziewiątej części książki.

Automatyczne uzyskiwanie parametrów objętościowych elementów z projektów CAD, które są niezbędne do przygotowania kosztorysów, odbywa się za pomocą narzędzi grupujących QTO (Quantity Take-Off). Narzędzia QTO działają poprzez grupowanie wszystkich obiektów projektu według specjalnych identyfikatorów elementów lub parametrów atrybutów elementów, przy użyciu specyfikacji i tabel utworzonych w bazie danych CAD.

QTO Quantity Take-Off: grupowanie danych projektu według atrybutów

QTO (Quantity Take-Off) w budownictwie to proces wyodrębniania ilości elementów wymaganych do realizacji projektu. W praktyce QTO jest często procesem częściowo ręcznym, obejmującym gromadzenie danych z różnych źródeł: dokumentów PDF, rysunków DWG i cyfrowych modeli CAD.

Podczas pracy z danymi pobranymi z baz danych CAD, proces QTO jest realizowany jako sekwencja operacji filtrowania, sortowania, grupowania i agregacji. Elementy modelu są wybierane zgodnie z parametrami klas, kategorii i typów, po czym ich atrybuty ilościowe - takie jak objętość, powierzchnia, długość lub ilość - są podsumowywane zgodnie z logiką obliczeń (rys. 5.2-2).



Rys. 5.2-2 Grupowanie i filtrowanie danych to najpopularniejsze funkcje stosowane w bazach danych i hurtowniach danych.

Proces QTO (filtrowanie i grupowanie) pozwala usystematyzować dane, utworzyć specyfikacje i przygotować informacje wejściowe do obliczania szacunków, zakupów i harmonogramów prac. Podstawą QTO jest klasyfikacja elementów według rodzaju mierzonych atrybutów. Dla każdego elementu lub grupy elementów wybierany jest odpowiedni parametr pomiaru ilościowego. Na przykład:

- **Atrybut długości** (krawężnik - w metrach)
- **Atrybut powierzchni** (płyta gipsowo-kartonowa - w metrach kwadratowych)
- **Atrybut objętości** (roboty betonowe - w metrach sześciennych)
- **Atrybut ilości** (okna - za sztukę)

Oprócz charakterystyk objętościowych generowanych matematycznie na podstawie geometrii, w obliczeniach po grupowaniu QTO często stosuje się współczynniki przekroczenia (rys. 5.2-12, np. 1,1, aby uwzględnić 10% na logistykę i instalację) - wartości korekcyjne uwzględniające straty, instalację, magazynowanie lub transport. Pozwala to dokładniej przewidzieć rzeczywiste zużycie materiałów i uniknąć zarówno niedoborów, jak i nadmiernych zapasów na placu budowy.

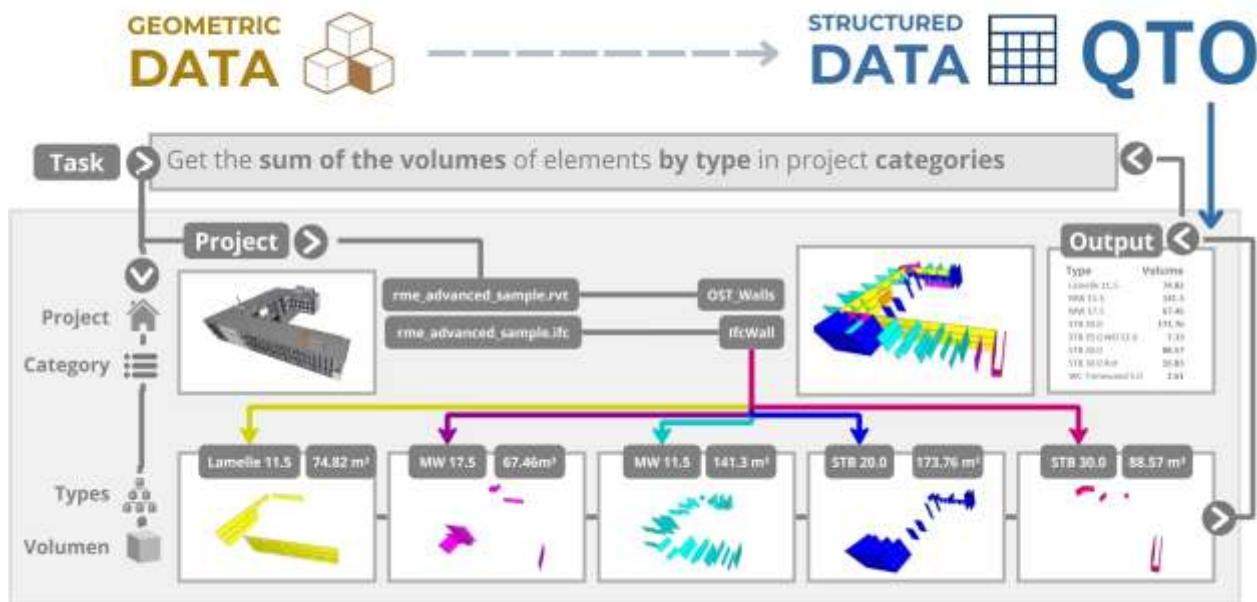
Zautomatyzowany proces odbioru ilościowego (QTO) jest niezbędny do tworzenia dokładnych obliczeń i szacunków, zmniejszania błędów ludzkich w procesach specyfikacji ilościowej i zapobiegania nadmiernemu lub niedostatecznemu zamawianiu materiałów.

Jako przykład procesu QTO rozważmy typowy przypadek, w którym konieczne jest przedstawienie z bazy danych CAD tabelarycznej specyfikacji objętości według typów elementów dla określonej kategorii, klas elementów. Pogrupujmy wszystkie elementy projektu według typu z kategorii ścian projektu CAD i podsumujmy atrybuty objętości dla każdego typu, aby przedstawić wynik jako tabelę objętości QTO (rys. 5.2-3).

W przykładzie typowego projektu CAD (rys. 5.2-3), wszystkie elementy kategorii ścian w bazie danych CAD są pogrupowane według typu ściany, np. "Lamelle 11.5", "MW 11.5" i "STB 20.0", i mają dobrze zdefiniowane atrybuty objętości reprezentowane w kostkach metrycznych.

Celem menedżera, który znajduje się na styku projektantów i specjalistów ds. obliczeń, jest uzyskanie zautomatyzowanej tabeli objętości według typu elementu w wybranej kategorii. Nie tylko dla konkretnego projektu, ale także w uniwersalnej formie mającej zastosowanie do innych projektów o podobnej strukturze modelu. Pozwala to na skalowalność podejścia i ponowne wykorzystanie danych bez powielania wysiłków.

Dawno minęły czasy, gdy doświadczeni projektanci i kosztorysanci uzbrojeni w linijkę, dokładnie mierzyli każdą linię na papierze lub planach PDF - tradycja, która nie zmieniła się przez ostatnie tysiąclecia. Wraz z rozwojem modelowania 3D, w którym geometria każdego elementu jest teraz bezpośrednio powiązana z automatycznie obliczanymi atrybutami objętościowymi, proces określania objętości i ilości QTO stał się zautomatyzowany.



Rys. 5.2-3 Uzyskanie atrybutów objętości i ilości QTO z projektu obejmuje grupowanie i filtrowanie elementów projektu.

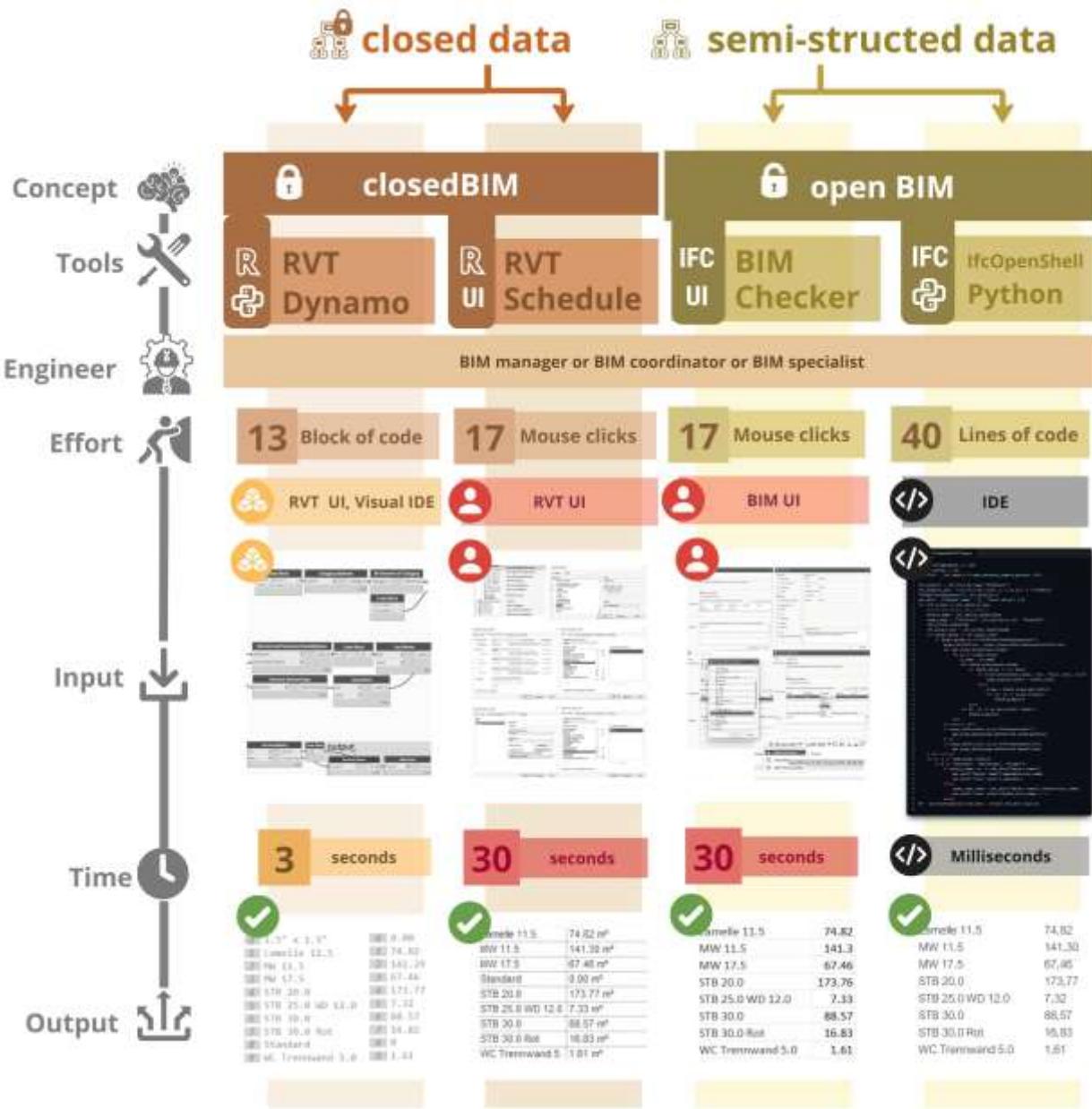
W naszym przykładzie zadaniem jest "wybranie kategorii ścian w projekcie, pogrupowanie wszystkich elementów według typu i przedstawienie informacji o atrybutach objętości w ustrukturyzowanym, tabelarycznym formacie", tak aby tabela ta mogła być używana przez dziesiątki innych specjalistów do obliczeń kosztorysów, logistyki, harmonogramów prac i innych przypadków biznesowych (rys. 6.1-3).

Ze względu na zamknięty charakter danych CAD nie każdy specjalista może dziś korzystać z bezpośredniego dostępu do bazy danych CAD (przyczyny i rozwiązania problemu dostępu zostały szczegółowo opisane w szóstej części książki). Dlatego wiele osób musi sięgać po specjalistyczne narzędzia BIM oparte na koncepcjach otwartego BIM i zamkniętego BIM [63]. Podczas pracy ze specjalistycznymi narzędziami BIM lub bezpośrednio w środowisku oprogramowania CAD, tabela z wynikami QTO (Quantity Take-Off) może być generowana na różne sposoby - w zależności od tego, czy używany jest interfejs ręczny, czy automatyzacja oprogramowania.

Na przykład, korzystając z interfejsu użytkownika oprogramowania CAD (BIM), wystarczy wykonać około 17 czynności (kliknąć przycisków), aby uzyskać gotową tabelę objętości (rys. 5.2-4). Użytkownik musi jednak dobrze rozumieć strukturę modelu i funkcje oprogramowania CAD (BIM).

Jeśli automatyzacja jest stosowana za pomocą kodu programu lub wtyczek i narzędzi API w programach CAD, liczba ręcznych kroków w celu uzyskania tabel objętości jest zmniejszona, ale konieczne będzie napisanie od 40 do 150 wierszy kodu, w zależności od używanej biblioteki lub narzędzia:

- **IfcOpSh (otwarty BIM)** lub **Dynamo IronPython (zamknięty BIM)** - pozwalają uzyskać tabelę QTO z formatu CAD lub programu CAD w zaledwie ~40 wierszach kodu.
- **IFC_js (open BIM)** - wymaga około 150 linii kodu, aby wyodrębnić obszerne atrybuty z modelu IFC.
- **Interfejs narzędzi CAD (BIM)** - pozwala uzyskać ten sam wynik ręcznie, za pomocą 17 kliknięć myszą.



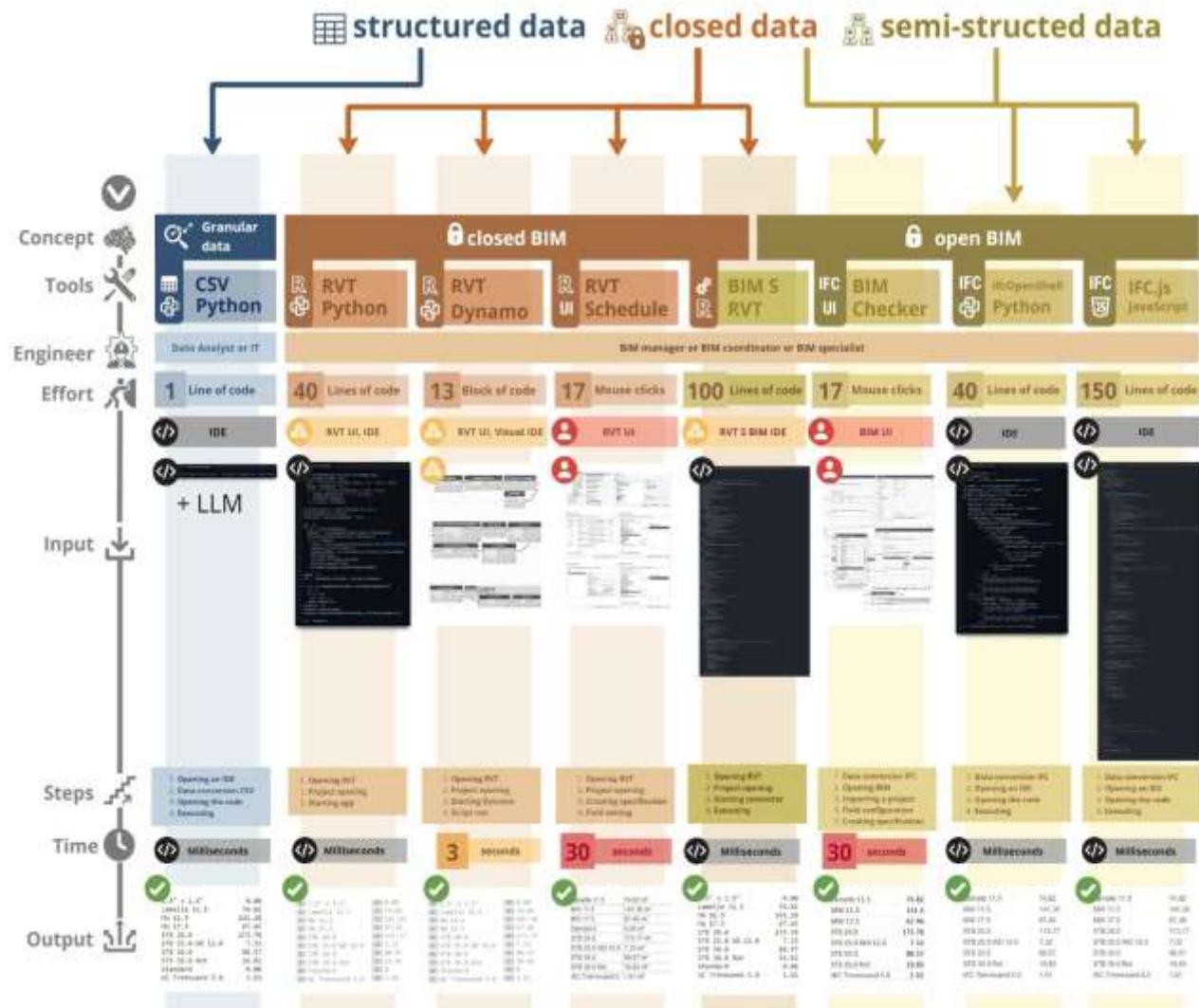
Rys. 5.2-4 Projektanci i menedżerowie CAD (BIM) używają od 40 do 150 wierszy kodu lub kilkunastu naciśnień klawiszy do tworzenia tabel QTO

Wynik jest taki sam - ustrukturyzowana tabela z atrybutami objętości dla grupy elementów. Jedyną różnicą jest koszt pracy i poziom wiedzy technicznej wymaganej przez użytkownika (rys. 5.2-4). Nowoczesne narzędzia, w stosunku do ręcznego gromadzenia volumenów, znacznie przyspieszają proces QTO i zmniejszają prawdopodobieństwo wystąpienia błędów. Pozwalają one na pobieranie danych bezpośrednio z modelu projektu, eliminując potrzebę ręcznego przeliczania objętości z rysunków, jak miało to miejsce w przeszłości.

Niezależnie od zastosowanej metody - otwartego BIM lub zamkniętego BIM - możliwe jest uzyskanie identycznej tabeli QTO z objętościami elementów projektu (rys. 5.2-4). Jednak podczas pracy z danymi

projektowymi w koncepcjach CAD - (BIM) użytkownicy polegają na specjalistycznych narzędziach i interfejsach API dostarczanych przez dostawców (rys. 3.2-13). Tworzy to dodatkowe warstwy zależności i wymaga nauki unikalnych schematów danych, jednocześnie ograniczając bezpośredni dostęp do danych.

Ze względu na zamknięty charakter danych CAD, uzyskanie tabel QTO i innych parametrów komplikuje automatyzację obliczeń i integrację z systemami zewnętrznymi. Korzystając z narzędzi bezpośredniego dostępu do baz danych i przenosząc dane CAD projektu za pomocą narzędzi inżynierii odwrotnej do otwartego formatu strukturalnej ramki danych (rys. 4.1-13), identyczną tabelę QTO można uzyskać za pomocą tylko jednej linii kodu (rys. 5.2-5 - wariant z danymi ziarnistymi).



Rys. 5.2-5 Różne narzędzia dają te same wyniki w postaci tabel atrybutów podmiotów projektu, ale przy różnych kosztach pracy.

W przypadku korzystania z otwartych danych strukturalnych z projektów CAD, jak wspomniano w rozdziale "Konwersja danych CAD (BIM) do postaci strukturalnej", proces grupowania, QTO, jest znacznie uproszczony.

Podejścia oparte na wykorzystaniu otwartych danych strukturalnych lub bezpośrednim dostępie do baz danych modeli CAD są wolne od ograniczeń marketingowych związanych ze skrótem BIM. Opierają się one

na sprawdzonych narzędziach od dawna stosowanych w innych branżach (rys. 7.3-10 Proces ETL).

Według badania McKinsey "Open Data: Unleash Innovation and Productivity with Streaming Information" [102] przeprowadzonego w 2013 r. [102] przeprowadzonego w 2013 r., wykorzystanie otwartych danych może stworzyć możliwości oszczędności od 30 do 50 miliardów dolarów rocznie w zakresie projektowania, inżynierii, zamówień i budowy obiektów energetycznych. Przekłada się to na 15-procentowe oszczędności w kosztach kapitału budowlanego.

Praca z otwartymi danymi strukturalnymi (granularnymi) upraszcza wyszukiwanie i przetwarzanie informacji, zmniejsza zależność od specjalistycznych platform BIM i otwiera drogę do automatyzacji bez konieczności korzystania z zastrzeżonych systemów lub parametrycznych i złożonych modeli danych z formatów CAD.

Automatyzacja QTO przy użyciu LLM i danych strukturalnych

Tłumaczenie nieustrukturyzowanych danych na ustrukturyzowaną formę znacznie poprawia wydajność różnych procesów: upraszcza przetwarzanie danych (rys. 4.1-1, rys. 4.1-2) i przyspiesza proces walidacji, czyniąc wymagania jasnymi i przejrzystymi, jak już omówiliśmy w poprzednich rozdziałach. Podobnie, tłumaczenie danych CAD (BIM) na ustrukturyzowaną formę otwartą (rys. 4.1-12, rys. 4.1-13) ułatwia proces grupowania atrybutów i proces QTO.

Tabela atrybutów QTO ma ustrukturyzowaną formę, więc podczas korzystania z ustrukturyzowanych danych CAD pracujemy z jednym modelem danych (Rys. 5.2-5), co eliminuje potrzebę konwersji i tłumaczenia modeli danych projektu i reguł grupowania do wspólnego mianownika. Pozwala nam to grupować dane według jednego lub więcej atrybutów za pomocą tylko jednej linii kodu. W przeciwieństwie do tego, w otwartym BIM i zamkniętym BIM, gdzie dane są przechowywane w formatach półstrukturalnych, parametrycznych lub zamkniętych, przetwarzanie wymaga dziesiątek, a nawet setek wierszy kodu i użycia interfejsu API do interakcji z geometrią i informacjami o atrybutach.

- 💡 Przykład grupowania projektu o strukturze QTO według jednego atrybutu. Zapytanie tekstowe na dowolnym czacie LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN lub dowolnym innym):

Mam CAD -project jako DataFrame - przefiltruj dane projektu, aby uzyskać elementy z parametrem "Type" zawierającym tylko wartość "Type 1" ↴

- Odpowiedź LLM jest bardzo prawdopodobna w postaci kodu Pythona przy użyciu Pandas :



Rys. 5.2-6 Jedna linia kodu napisana za pomocą LLM , pozwala na pogrupowanie całego projektu CAD według atrybutu "Type" i uzyskanie żądanej grupy elementów.

Dzięki prostej strukturze 2D DataFrame nie musimy wyjaśniać schematu i modelu danych LLM, co skraca etapy interpretacji i przyspiesza tworzenie ostatecznych rozwiązań. Wcześniej pisanie nawet prostego kodu wymagało nauki języków programowania, ale obecnie nowoczesne modele językowe (LLM) pozwalają nam automatycznie tłumaczyć logikę procesu na kod podczas pracy z danymi strukturalnymi za pomocą zapytań tekstowych.

Automatyzacja i modele językowe LLM mogą całkowicie wyeliminować konieczność nauki języków programowania lub narzędzi BIM przez profesjonalistów pracujących z grupowaniem i przetwarzaniem danych CAD (BIM), zapewniając możliwość rozwiązywania problemów za pomocą zapytań tekstowych.

To samo zapytanie - grupujące wszystkie elementy projektu z kategorii "ściany" i obliczające objętości dla każdego typu (rys. 5.2-5) - które w środowisku CAD (BIM) wymaga 17 kliknięć w interfejsie lub napisania 40 linii kodu, w otwartych narzędziach do przetwarzania danych (np. SQL lub Pandas) wygląda jak proste i intuicyjne zapytanie:

- Z pojedynczą linią w Pandas:

```
df[df['Category'].isin(['OST_Walls'])].groupby('Type')['Volume'].sum()
```

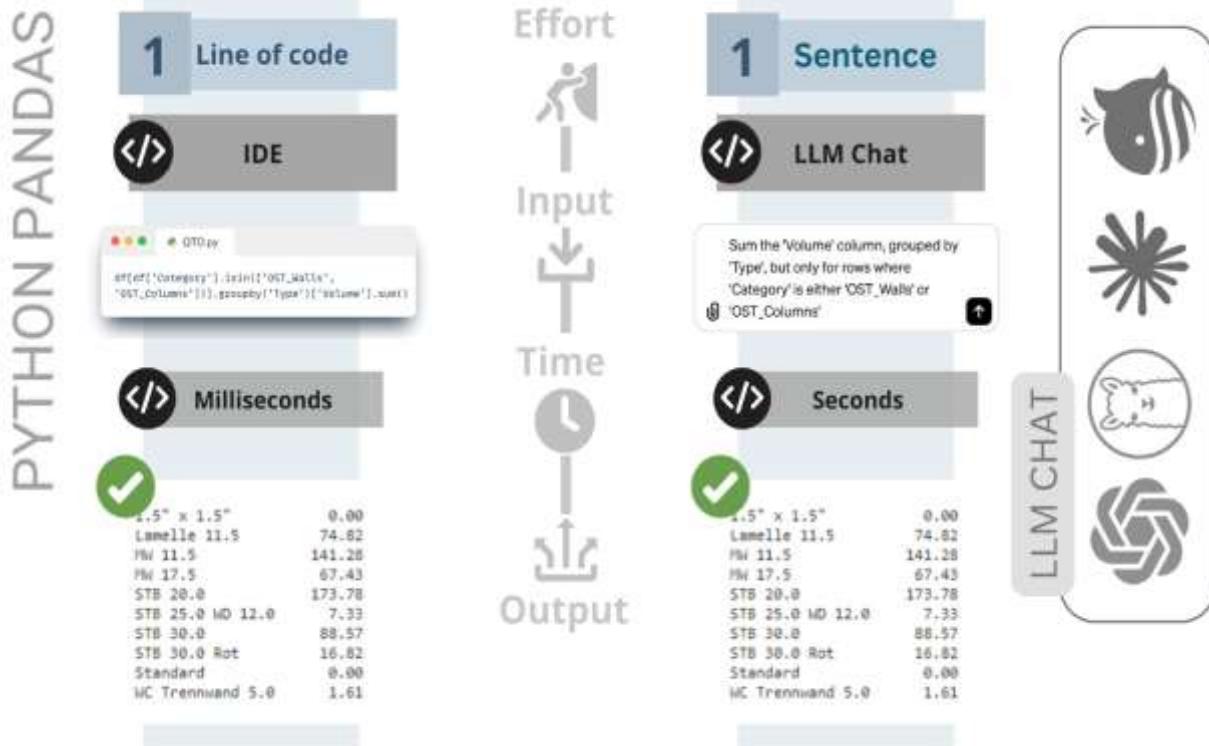
Dekodowanie kodu: pobierz z df (DataFrame) elementy, których atrybut-kolumna "Category", ma wartości "OST_Walls", pogrupuj wszystkie uzyskane elementy według atrybutu-kolumny "Type" i zsumuj dla uzyskanej grupy elementów atrybut "Volume".

- Grupowanie ustrukturyzowanego projektu uzyskanego z CAD przy użyciu SQL:

```
SELECT Type, SUM(Volume) AS TotalVolume
FROM elements
WHERE Category = 'OST_Walls'
GROUP BY Type;
```

- Z pomocą LLM możemy napisać żądanie grupowania do bazy danych projektu jako prosty odnośnik tekstowy - monit (rys. 5.2-7):

W przypadku ramki danych projektu pogrupuj elementy według parametru "Type", ale tylko dla elementów z parametrem "Category" równym "OST_Walls" lub "OST_Columns" i podsumuj parametr kolumny "Volume" dla wynikowej. ↵

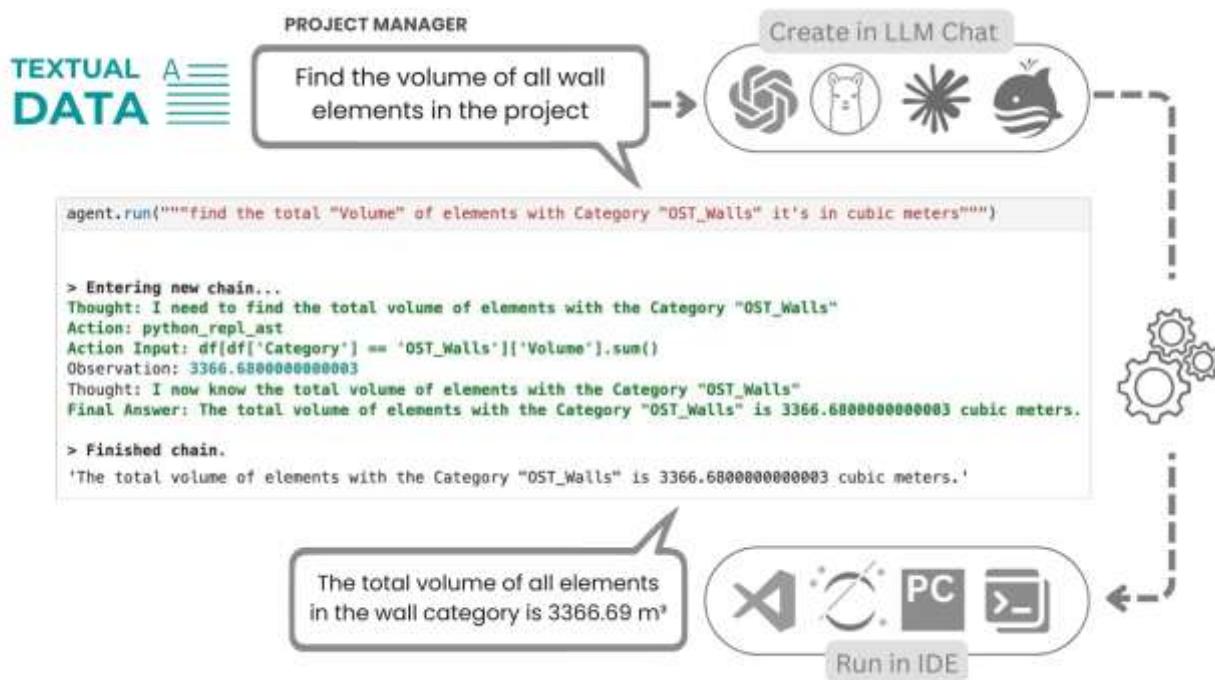


Rys. 5.2-7 Korzystanie z SQL, Pandas i LLM automatyzacja przetwarzania danych jest teraz możliwa za pomocą kilku wierszy kodu i zapytań tekstowych.

Uzyskanie QTO z danych CAD przy użyciu narzędzi LLM (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok), radykalnie zmienia tradycyjne metody wyodrębniania informacji atrybutowych, danych ilościowych i objętościowych dla poszczególnych obiektów i grup obiektów.

Teraz nawet kierownicy projektów, specjaliści ds. kosztorysowania lub logistyki, którzy nie mają dogłębnej wiedzy na temat projektowania i nie posiadają specjalistycznego oprogramowania CAD - (BIM-), mając dostęp do bazy danych CAD, mogą uzyskać całkowitą objętość elementów kategorii ścian lub innych obiektów w ciągu kilku sekund, po prostu pisząc lub dyktując zapytanie.

W zapytaniach tekstowych (rys. 5.2-8) agent LLM modelu przetwarza żądanie użytkownika dotyczące zastosowania określonej funkcji do jednego lub więcej parametrów - kolumn tabeli. W rezultacie użytkownik komunikujący się z LLM otrzymuje albo nowy parametr kolumny z nowymi wartościami, albo jedną konkretną wartość po zgrupowaniu



Rys. 5.2-8 LLM model, pracujący z danymi strukturalnymi, rozumie z kontekstu zapytania tekstuowego, o jakie grupy i atrybuty pyta użytkownik.

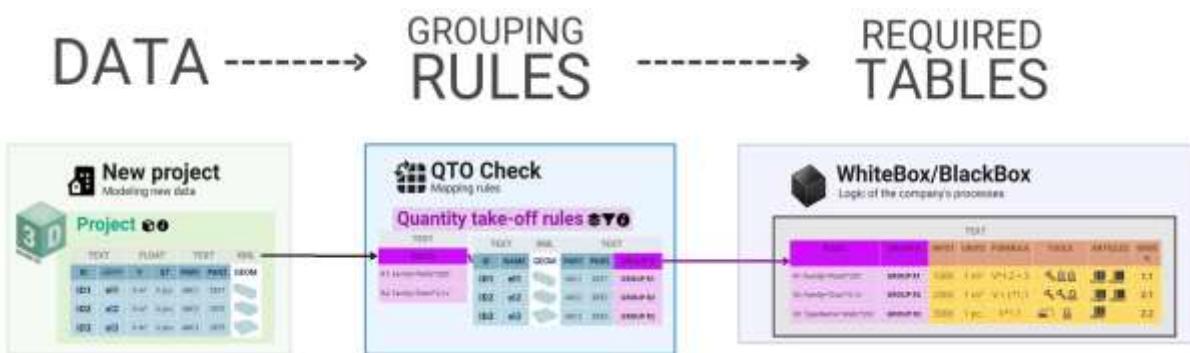
Jeśli konieczne jest uzyskanie ilości tylko dla jednej grupy elementów, wystarczy wykonać proste zapytanie QTO (Rys. 5.2-7) na danych modelu CAD. Jednak przy obliczaniu budżetu lub szacunków dla całego projektu składającego się z wielu grup elementów często konieczne jest wyodrębnienie ilości dla wszystkich typów elementów (klas0, gdzie każda kategoria elementów jest przetwarzana osobno - z grupowaniem według odpowiednich atrybutów.

W praktyce kosztorysantów i rzeczników stosuje się indywidualne zasady grupowania i obliczania dla różnych typów obiektów. Na przykład okna są zwykle grupowane według pięter lub stref (parametr grupowania - atrybut Poziom, Pomieszczenia), a ściany - według materiału lub typu konstrukcji (parametr Materiał, Typ). Aby zautomatyzować proces grupowania, takie reguły są z góry opisane w postaci tabel reguł grupowania. Tabele te działają jako szablony konfiguracji, które definiują, które atrybuty powinny być

używane w obliczeniach dla każdej grupy elementów w projekcie.

QTO obliczanie całego projektu przy użyciu reguł grupowych z arkusza kalkulacyjnego Excel

W rzeczywistych projektach budowlanych często konieczne jest przeprowadzenie agregacji według kilku atrybutów jednocześnie w ramach jednej grupy elementów. Przykładowo, podczas pracy z kategorią "Windows" (gdzie atrybut Category zawiera wartości takie jak OST_Windows lub IfcWindows), elementy mogą być grupowane nie tylko według typu - na przykład według wartości w polu Type Name lub Type - ale także według dodatkowych cech, takich jak poziom przewodności cieplnej określony w odpowiednim atrybucie. Takie wielowymiarowe grupowanie pozwala na uzyskanie dokładniejszych wyników dla danej grupy. Podobnie, podczas obliczania kategorii ścian lub podłóg, dowolne kombinacje atrybutów - takich jak materiał, poziom, podłoga, odporność ognowa i inne parametry - mogą być używane jako filtry lub kryteria grupowania (Rys. 5.2-9).



Rys. 5.2-9 Dla każdej grupy lub kategorii podmiotów w projekcie istnieje inna formuła grupowania składająca się z jednego lub więcej kryteriów.

Proces definiowania takich reguł grupowania jest podobny do procesu tworzenia wymagań dotyczących danych opisanego w rozdziale "Tworzenie wymagań i sprawdzanie jakości danych" (Rys. 4.4-5), gdzie szczegółowo omówiliśmy pracę z modelami danych. Te reguły grupowania i obliczania zapewniają dokładność i trafność wyników, aby automatycznie obliczyć łączne atrybuty ilości lub objętości kategorii encji, biorąc pod uwagę wszystkie niezbędne warunki, które muszą być uwzględnione w obliczeniach i obliczeniach.

- » Poniższa próbka kodu filtuje tabelę projektów tak, aby wynikowy zestaw danych zawierał tylko encje, w których kolumna atrybutu "Category" zawiera wartości "OST_Windows" lub "IfcWindows" i jednocześnie kolumna atrybutu "Type" zawiera wartość "Type 1":

Mam projekt DataFrame - przefiltruj dane tak, aby w zbiorze danych pozostały tylko elementy, których atrybut "Category" zawiera wartości "OST_Windows" lub "IfcWindows" i jednocześnie atrybut Type zawiera wartość "Type 1" ↴

💡 Odpowiedź LLM :



```
group.py
1 df[(df['Category'].isin(['OST_Windows', 'IfcWindows'])) & (df['Type'].str.contains("Type 1"))]
```

The screenshot shows a Python code editor window titled "group.py". Inside the editor, there is a single line of code: `df[(df['Category'].isin(['OST_Windows', 'IfcWindows'])) & (df['Type'].str.contains("Type 1"))]`. Above the editor, there is a toolbar with icons for "Create in LLM Chat" and several AI-related icons. Below the editor, there are icons for "Run in IDE" and other development tools like VS Code, PyCharm, and Jupyter Notebook.

Rys. 5.2-10 Pojedyncza linia kodu, podobna do formuły programu Excel , umożliwia grupowanie wszystkich jednostek projektu według kilku atrybutów.

Kod wynikowy (rys. 5.2-10) po przetłumaczeniu danych CAD w ustrukturyzowanych otwartych formatach (rys. 4.1-13) można uruchomić w jednym z popularnych IDE (zintegrowanych środowisk programistycznych), o których wspomnialiśmy powyżej w trybie offline: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse z wtyczką PyDev, Thonny, Wing IDE, IntelliJ IDEA z wtyczką Python, JupyterLab lub popularnych narzędziach online: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

💡 Aby pobrać jednostki projektu w formularzu QTO DataFrame w kategorii "Windows" tylko z określoną wartością przewodności cieplnej, możemy użyć następującego zapytania do LLM:

Mam projekt DataFrame - przefiltruj dane tak, aby w zbiorze danych pozostały tylko rekordy z "Category" zawierające wartości "OST_Windows" lub "IfcWindows", a jednocześnie kolumna ThermalConductivity powinna mieć wartość 0, ↴

💡 Odpowiedź LLM :



```
group.py
1 df[(df['Category'].isin(['OST_Windows', 'IfcWindows'])) & (df['ThermalConductivity'] == 0.5)]
```

The screenshot shows a Python code editor window titled "group.py". Inside the editor, there is a line of code: `df[(df['Category'].isin(['OST_Windows', 'IfcWindows'])) & (df['ThermalConductivity'] == 0.5)]`. Above the editor, there is a toolbar with icons for "Create in LLM Chat" and several AI-related icons. Below the editor, there are icons for "Run in IDE" and other development tools like VS Code, PyCharm, and Jupyter Notebook.

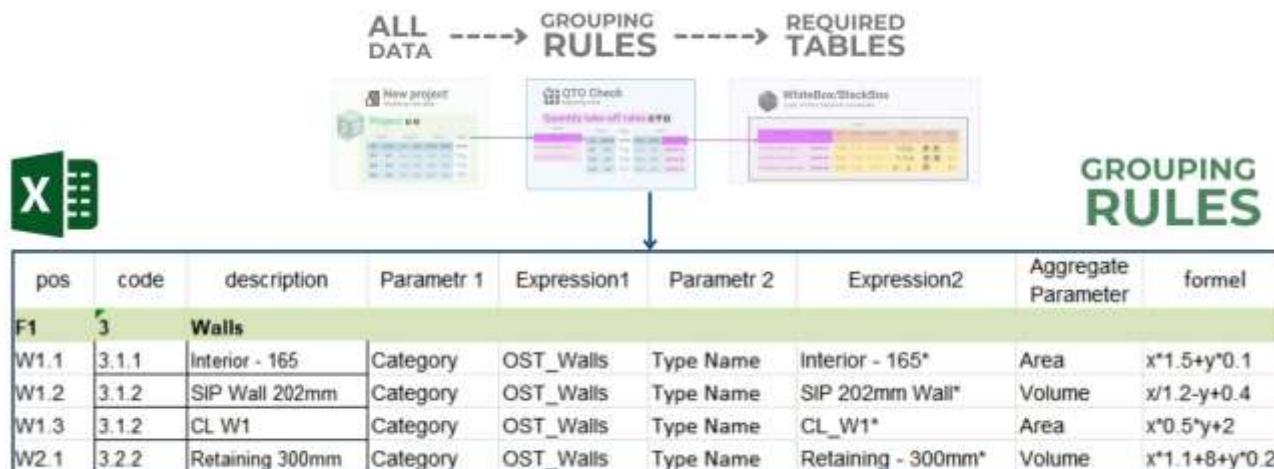
Rys. 5.2-11 Niezwykle prosty język zapytań Pandas Python umożliwia uruchamianie QTO dla dowolnej liczby projektów jednocześnie.

W odpowiedzi otrzymanej od LLM (rys. 5.2-11) warunek logiczny "&" jest używany do połączenia dwóch kryteriów: wartości przewodności cieplnej i przynależności do jednej z dwóch kategorii. Metoda "isin" sprawdza, czy wartość atrybutu kolumny "Category" znajduje się na podanej liście.

W projektach z dużą liczbą grup elementów, z różną logiką grupowania - dla każdej kategorii podmiotów projektu (np.: okna, drzwi, płyty) należy ustalić indywidualne zasady grupowania, które mogą obejmować dodatkowe współczynniki lub całkowite formuły obliczania atrybutów. Formuły te (Rys. 5.2-12 "formuła" atrybutu, np. x-wartość ilości i y-objętość grupy) i współczynniki uwzględniają na przykład unikalne cechy każdej grupy:

- % dodatków do ilości materiału w celu uwzględnienia przekroczeń
- ustalona dodatkowa ilość materiału
- korekty związane z możliwym ryzykiem i błędami obliczeniowymi w postaci wzorów

Po sformułowaniu reguł filtrowania i grupowania w postaci formuł parametrów dla każdej kategorii pozycji, mogą one być przechowywane jako tabela wiersz po wierszu - na przykład w formacie Excel (Rys. 5.2-12). Dzięki przechowywaniu tych reguł w ustrukturyzowanej formie, proces wyodrębniania, filtrowania i grupowania danych projektu może być w pełni zautomatyzowany. Zamiast ręcznie pisać wiele oddzielnych zapytań, system po prostu odczytuje tabelę parametrów i stosuje odpowiednie reguły do modelu (ogólna ramka danych projektu (rys. 4.1-13)), generując ostateczne tabele QTO dla każdej kategorii elementów projektu.



Rys. 5.2-12 Tabela grupowania atrybutów QTO ustanawia zasady grupowania elementów projektu, zapewniając dokładnąłączną liczbę i zakres dla każdej kategorii.

Zebrane reguły pozwolą na pogrupowanie całego projektu i wykonanie wszystkich niezbędnych obliczeń, w tym dostosowanie atrybutów objętości. W rezultacie objętości są doprowadzane do "rzeczywistej objętości", która jest używana do obliczeń i obliczeń, a nie tych, które były pierwotnie na etapie projektowania w modelu CAD.

W procesie automatycznego tworzenia tabel objętości QTO dla całego projektu aplikacja musi przejść przez wszystkie kategorie tabeli reguł grupowania, pobrać atrybuty grupowania, pogrupować według nich wszystkie elementy projektu i zagregować atrybut objętości dla tej grupy, dodatkowo mnożąc go przez współczynnik lub współczynnik rafinacji.

Poprośmy LLM o napisanie dla nas kodu dla takiego rozwiązania, w którym kod musiałby załadować dwie tabele - tabelę reguł grupowania (Rys. 5.2-12) i tabelę danych z samego projektu (Rys. 4.1-13), a następnie zastosować reguły grupowania, pogrupować elementy zgodnie z podanymi regułami, obliczyć zagregowane wartości i zapisać wyniki do nowego pliku Excel.

✉ Wyślij prośbę tekstową na czat LLM:

Potrzebuję kodu, który odczyta dane projektu z pliku "basic_sample_project.xlsx", a następnie reguły z "Grouping_rules_QTO.xlsx", za pomocą których pogrupuje wszystkie dane według "Parameter 1" i "Parameter 2", zagraruje "Aggregate Parameter", przefiltruje według "Expression2", wykona obliczenia z "Formel1" i zapisze tabelę QTO do "QTQ_table2.xlsx" ↵

2 Odpowiedź LLM :



```

1 import pandas as pd
2
3 # Reading CSV and Excel files with project data and grouping rules respectively
4 df = pd.read_csv('../input/rac_basic_sample_project.xlsx', low_memory=False)
5 df_qto = pd.read_excel('../input//Grouping_rules_QTO.xlsx', header=0, skiprows=[1])
6
7 # Cleaning numerical columns in the dataframe by removing non-numeric characters
8 numeric_cols = ['Area', 'Volume', 'Depth', 'Width', 'Length', 'Perimeter']
9 df[numeric_cols] = df[numeric_cols]
10 ].apply(lambda col: pd.to_numeric(col.astype(str).str.extract('(\d+\.\d*)')[0]).fillna(0))
11
12 # Define a function to process each row in the grouping rules dataframe
13 def process_row(row):
14     # Group by specified parameters and calculate sum and count for the aggregate parameter
15     group = df.groupby([row["Parameter 1"], row["Parameter 2"]])
16     ][row["Aggregate Parameter"]].agg(['sum', 'count'])
17     group = group[group.index.get_level_values(1).str.contains(row["Expression2"], regex=True)]
18
19     # Perform calculation defined in the rules
20     group['After_Calculation'] = eval(row["formel"], {'x': group['sum'], 'y': group['count']})
21
22     return group.reset_index()
23
24 # Apply the function to each row in the grouping rules dataframe and concatenate the results
25 dfgroups = pd.concat([process_row(row) for _, row in df_qto.iterrows()]).reset_index(drop=True)
26
27 # Save the processed data to an Excel file using the OpenPyXL engine
28 dfgroups.to_excel('QTO_table2.xlsx', engine='openpyxl', index=False)

```

Rys. 5.2-13 Kod Python Pandas przetwarza dane projektowe z projektu CAD , grupując je zgodnie z regułami zdefiniowanymi w programie Excel .

Końcowym wynikiem wykonania kodu (Rys. 5.2-13) będzie tabela grupy encji, która zawiera nie tylko uogólnione atrybuty objętości z oryginalnego modelu CAD- (BIM-), ale także nowy rzeczywisty atrybut objętości, który uwzględnia wszystkie wymagania dotyczące prawidłowego tworzenia obliczeń i szacunków (Przykład Rys. 5.2-14).

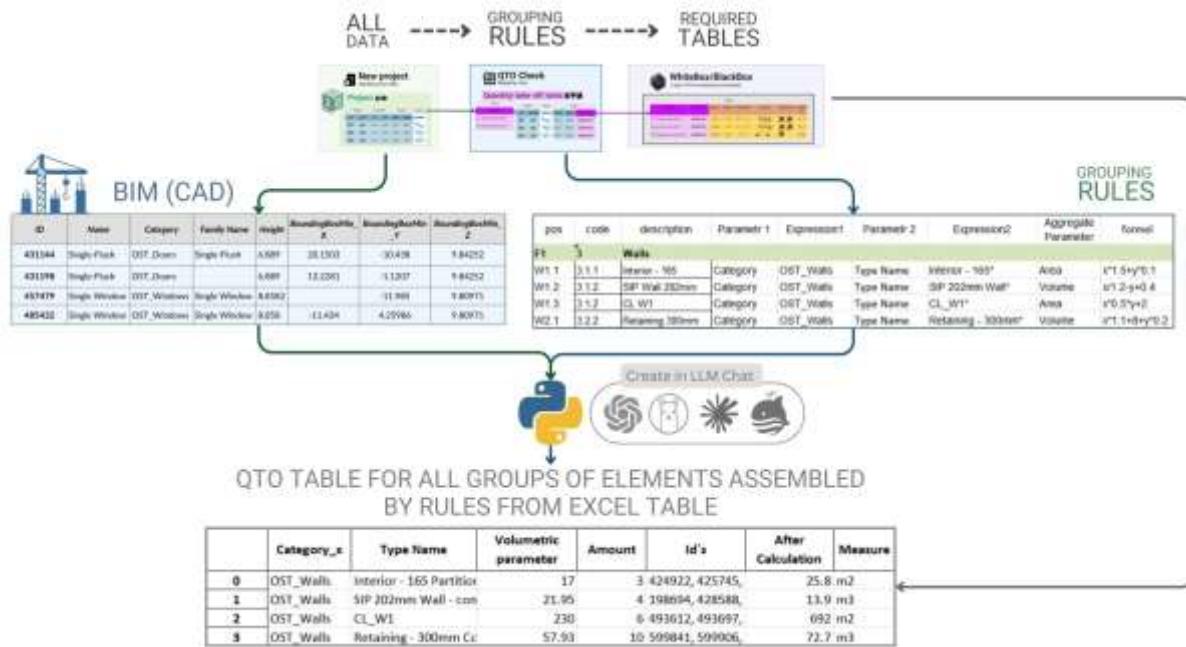
**QTO TABLE FOR ALL GROUPS OF ELEMENTS
ASSEMBLED BY RULES FROM EXCEL TABLE**

	Category_x	Type Name	Volumetric parameter	Amount	Id's	After Calculation	Measure
0	OST_Walls	Interior - 165 Partition	17	3 424922, 425745,		25.8 m2	
1	OST_Walls	SIP 202mm Wall - con	21.95	4 198694, 428588,		13.9 m3	
2	OST_Walls	CL_W1	230	6 493612, 493697,		692 m2	
3	OST_Walls	Retaining - 300mm Cc	57.93	10 599841, 599906,		72.7 m3	

Rys. 5.2-14 Atrybut "After Calculation" jest dodawany do tabeli podsumowującej po wykonaniu kodu, który automatycznie obliczy rzeczywistą objętość.

Wynikowy kod (rys. 5.2-13) można uruchomić w jednym z popularnych IDE (o których wspomnialiśmy

powyżej) i zastosować do dowolnej liczby istniejących lub nowych projektów przychodzących (RVT, IFC, DWG, NWS, DGN itp.), niezależnie od tego, czy jest to kilka projektów, czy może setki projektów w różnych formatach w ustrukturowanej formie (rys. 5.2-15).



Rys. 5.2-15 Automatyczny proces grupowania danych budowlanych łączy dane BIM (CAD) z tabelami QTO za pomocą reguł z arkusza kalkulacyjnego Excel.

Dostosowany i sparametryzowany proces gromadzenia danych wolumetrycznych (Rys. 5.2-15) umożliwia w pełni zautomatyzowane gromadzenie danych dotyczących atrybutów ilościowych i objętości elementów projektu w celu dalszej pracy z nimi, w tym szacowania kosztów, logistyki, harmonogramów prac i obliczania śladu węglowego oraz innych zadań analitycznych.

Po zapoznaniu się z narzędziami, które pozwalają nam łatwo organizować i grupować grupy elementów projektu według określonych atrybutów, jesteśmy teraz gotowi do integracji pogrupowanych i przefiltrowanych projektów z różnymi obliczeniami firmy i scenariuszami biznesowymi.



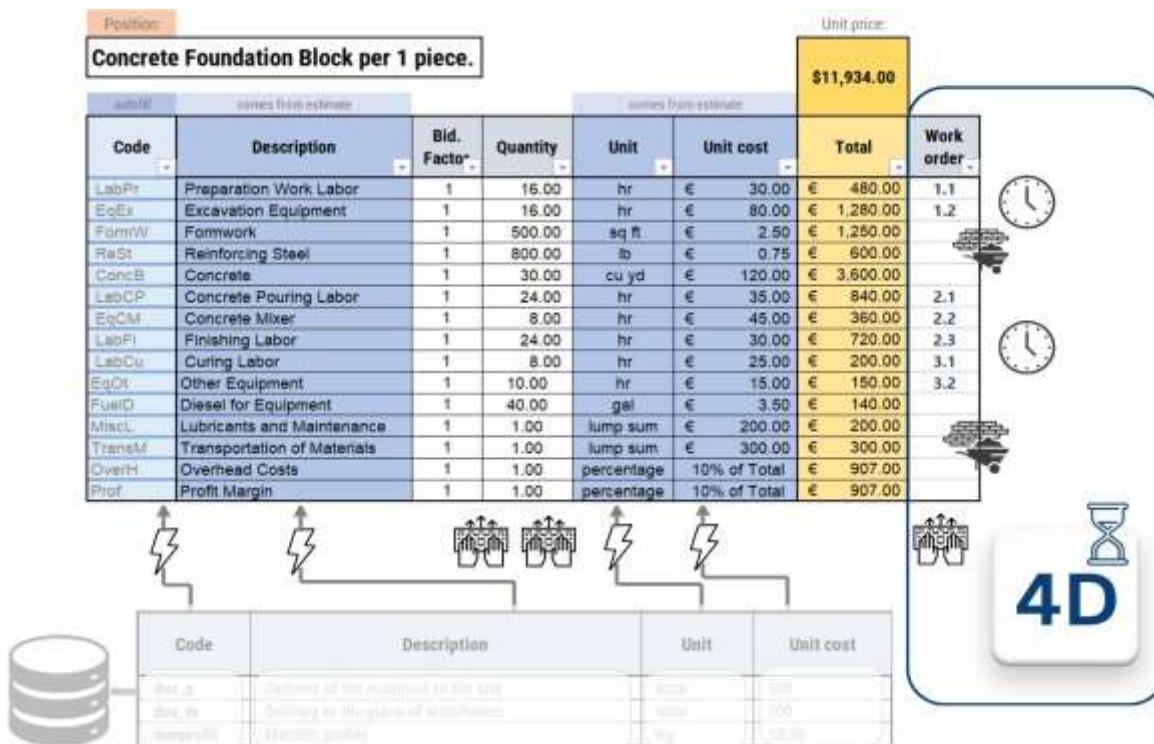
ROZDZIAŁ 4.3.

4D, 6D -8D I OBLICZENIA EMISJI DWUTLENKU WĘGLA CO₂

Model 4D: uwzględnienie czasu w szacunkach konstrukcyjnych

Oprócz kalkulacji kosztów, jednym z kluczowych zastosowań danych projektowych w budownictwie jest określenie parametrów czasowych - zarówno dla poszczególnych operacji budowlanych, jak i dla całego projektu. Metoda szacowania oparta na zasobach i związana z nią baza danych obliczeniowych, która została szczegółowo omówiona w poprzednim rozdziale "Obliczenia i szacunki dla projektów budowlanych", są często wykorzystywane jako podstawa do automatycznego obliczania czasu i tworzenia harmonogramu wykonywania prac.

Podejście oparte na zasobach uwzględnia nie tylko koszty materiałowe, ale także zasoby czasowe. W kalkulacji kosztów każdemu procesowi można przypisać atrybut zlecenia pracy (Rys. 5.3-1 - Parametr zlecenia pracy) oraz ilość czasu i koszt związany z wykonaniem tego procesu. Parametry te są szczególnie ważne przy opisywaniu operacji, które nie mają ustalonej ceny rynkowej i nie można ich bezpośrednio kupić - takich jak wykorzystanie sprzętu budowlanego, zatrudnienie robotników lub procesy logistyczne (które są zwykle wyrażane w godzinach). W takich przypadkach koszt nie jest określany przez dział zamówień, ale bezpośrednio przez firmę wdrażającą w oparciu o wewnętrzne normy lub wskaźniki produkcji (Rys. 5.3-1).



Rys. 5.3-1 Obliczenia pracy w metodzie szacowania opartej na zasobach obejmują harmonogram godzin pracy.

W związku z tym kalkulacje na poziomie kosztów obejmują nie tylko koszty paliwa i materiałów (koszty zakupu), ale także czas spędzony przez kierowców, techników i pracowników pomocniczych na miejscu. W

podanym przykładzie (rys. 5.3-1) tabela kosztów stanowi kalkulację kosztów montażu bloku fundamentowego, w tym etapów składowych prac, takich jak przygotowanie, montaż ramy i wylewanie betonu, a także wymaganych materiałów i robocizny. Jednak poszczególne operacje, takie jak prace przygotowawcze, mogą nie wiązać się z żadnymi kosztami materiałymi, ale mogą obejmować znaczne tymczasowe koszty robocizny wyrażone w roboczogodzinach.

Aby zaplanować sekwencję prac (dla harmonogramu prac) na placu budowy, atrybut "Zlecenie pracy" jest ręcznie dodawany do tabeli obliczeniowej (rys. 5.3-1). Atrybut ten jest określany w dodatkowej kolumnie tylko dla pozycji, których jednostka miary jest wyrażona w czasie (godzina, dzień). Atrybut ten jest dodatkiem do kodu pracy, opisu, ilości, jednostki miary (parametr "Jednostka") i kosztów. Sekwencja numeryczna (parametr "Kolejność prac") czynności pozwala ustalić kolejność wykonywania zadań na placu budowy i wykorzystać ją do tworzenia harmonogramu.

Harmonogram budowy i jego automatyzacja w oparciu o dane kosztorysowe

Harmonogram budowy to wizualna reprezentacja planu prac i procesów, które mają zostać przeprowadzone w ramach realizacji projektu. Jest on tworzony na podstawie szczegółowych obliczeń zasobów (Rys. 5.3-1), gdzie każde zadanie-zadanie jest zaplanowane, oprócz kosztu zasobów, według czasu i kolejności.

W przeciwieństwie do podejścia uśredniającego, w którym szacunki czasu opierają się na typowych godzinach instalacji materiałów lub sprzętu, w metodzie opartej na zasobach planowanie opiera się na rzeczywistych danych w kosztorysie. Każda pozycja szacunku związana z kosztami pracy opiera się na zastosowanym kalendarzu, który uwzględnia rzeczywiste warunki wykorzystania zasobów w okresie roboczym. Dostosowanie godzin produktywnych za pomocą współczynników na poziomie kosztorysu (rys. 5.3-1 parametr "Bid. Factor") umożliwia uwzględnienie różnic w produktywności i sezonowych osobliwości wpływających na harmonogram pracy.

Aby określić daty rozpoczęcia i zakończenia procesu dla harmonogramu budowy na wykresie Gantta, bierzemy wartości atrybutu ilości czasu dla każdej pozycji z kosztorysu bloku fundamentowego i mnożymy je przez liczbę bloków (w tym przypadku liczbę betonowych bloków fundamentowych). Obliczenia te dają czas trwania każdego zadania. Następnie wykreślamy te czasy trwania na osi czasu, zaczynając od daty rozpoczęcia projektu, aby utworzyć wykres, a wynikiem jest wizualna reprezentacja pokazująca, kiedy każde zadanie powinno się rozpocząć i zakończyć. Parametr "Kolejność pracy" dla procesów pozwala nam dodatkowo zrozumieć, czy proces pracy przebiega równolegle ("Kolejność pracy" np. 1.1-1.1) czy sekwencyjnie (1.1-1.2).

Wykres Gantta to graficzne narzędzie do planowania i zarządzania projektami, które przedstawia zadania jako poziome paski na osi czasu. Każdy słupek reprezentuje czas trwania zadania, jego początek i koniec.

Harmonogram lub wykres Gantta pomaga kierownikom projektu i pracownikom jasno zrozumieć, kiedy i w jakiej kolejności należy wykonać poszczególne etapy budowy, zapewniając efektywne wykorzystanie zasobów i dotrzymanie terminów.

Przedstawmy harmonogram instalacji trzech betonowych bloków fundamentowych przy użyciu obliczeń z powyższej tabeli. Korzystając z tabeli kosztów (rys. 5.3-1) z powyższego przykładu, poproś LLM o

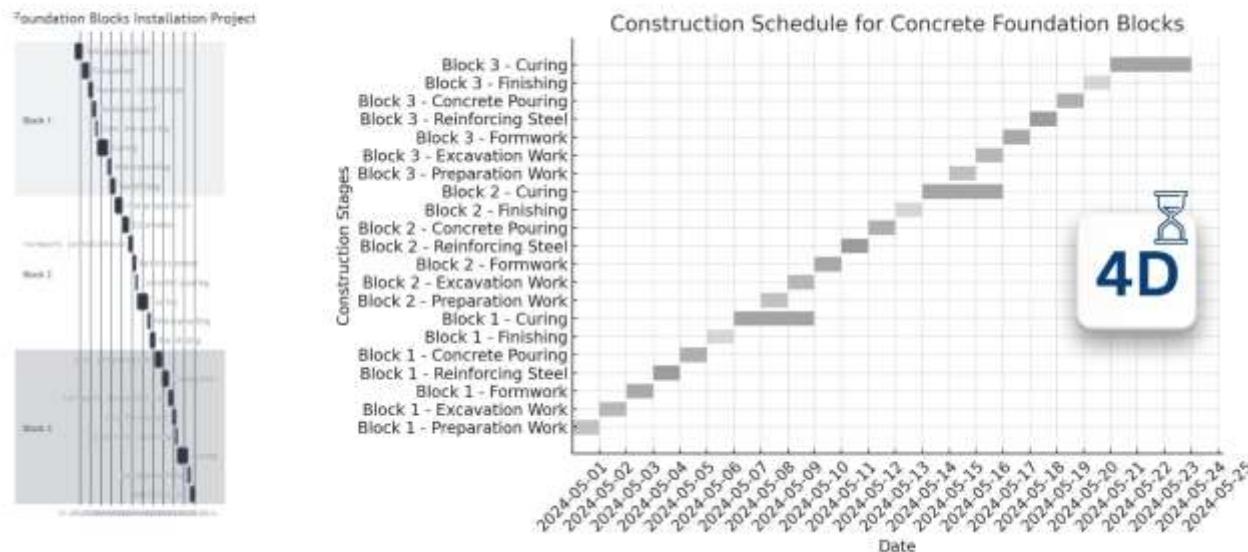
zaplanowanie montażu 3 elementów fundamentowych, np. na pierwszego maja 2024 roku.

Aby wysłać kosztorys do LLM, możemy przesłać tabelę kosztorysu w formacie XLSX lub po prostu wstawić zrzut ekranu z obrazem kosztorysu w formacie JPEG bezpośrednio do czatu LLM (rys. 5.3-1). LLM samodzielnie znajdzie bibliotekę do wizualizacji obrazu tabeli i, mnożąc atrybuty czasowe działań z tabeli przez ich objętość, doda wszystkie dane do wykresu.

- ⌚ Wyślij prośbę tekstową do LLM:

Utwórz harmonogram prac w formie wykresu Gantta, wykorzystując wartości z tabeli powiązanej z czasem (rys. 5.3-1 jako JPEG w załączniku) dla instalacji 3 bloków fundamentowych. Prace dla każdego bloku powinny być wykonywane sekwencyjnie. Początek prac należy ustalić na dzień 01/05/2024 r. ↳

- ⌚ Odpowiedź LLM



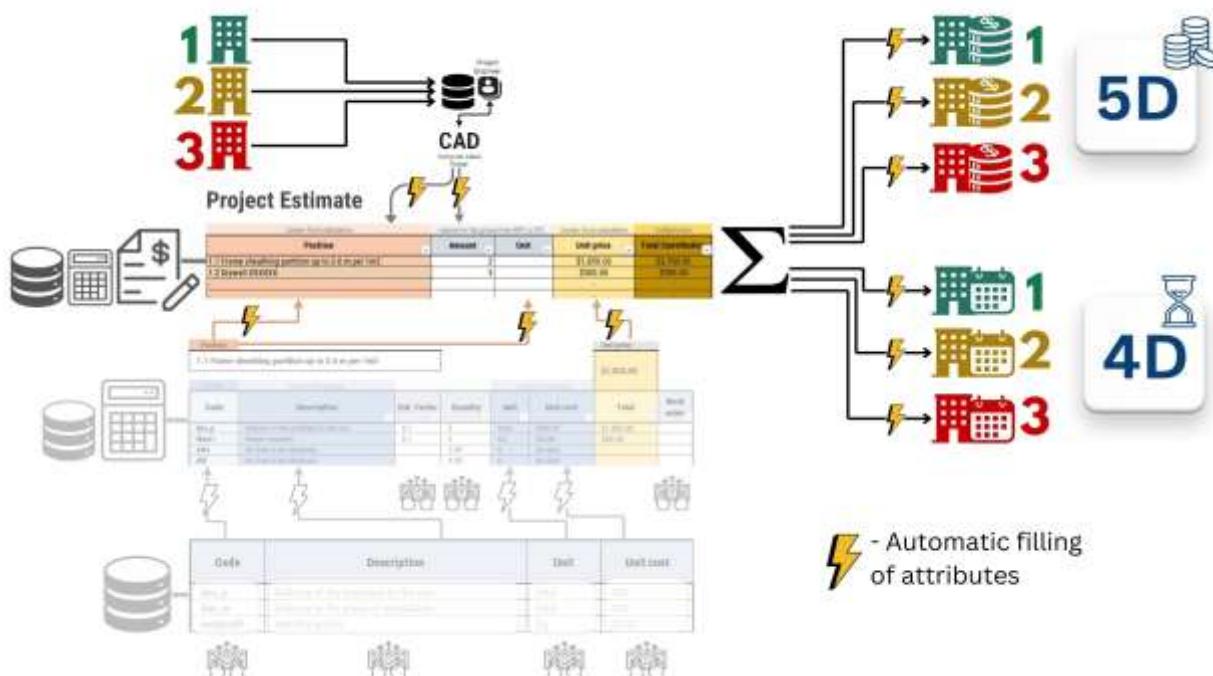
Rys. 5.3-2 Automatycznie generowany przez kilka LLM Wykres Gantta pokazuje etapy budowy trzech betonowych bloków, zgodnie z warunkami z monitu.

Wynikowy wykres (rys. 5.3-2) jest wykresem czasowym, w którym każdy poziomy słupek odpowiada pewnemu etapowi pracy nad blokiem fundamentowym i pokazuje kolejność operacji (parametr "Kolejność prac"), takich jak przygotowanie, wykopy, montaż szalunku, zbrojenie, wylewanie betonu i wykończenie, tj. te procesy, które mają wypełnione parametry czasowe i kolejność w obliczeniach.

Taki harmonogram (rys. 5.3-2) nie uwzględnia ograniczeń związanych z dniami roboczymi, zmianami lub normami czasu pracy, ale jest przeznaczony wyłącznie do koncepcyjnej wizualizacji procesu. Dokładny harmonogram, który będzie odzwierciedlał równoległość pracy, można uzupełnić odpowiednimi promesami lub dodatkowymi instrukcjami na czacie.

Korzystając z pojedynczego kosztorysu (rys. 5.3-1), dzięki atrybutom objętości z geometrii 3D, możliwe jest automatyczne oszacowanie zarówno kosztów projektu za pomocą zautomatyzowanych szacunków, jak i

jednoczesne obliczenie charakterystyki czasowej grup w postaci tabel lub wykresów dla różnych wariantów projektu (rys. 5.3-3).



Rys. 5.3-3 Automatyczna kalkulacja, umożliwia natychmiastowe i automatyczne prognozowanie kosztów i czasu dla różnych opcji projektu.

Nowoczesne modułowe systemy ERP (rys. 5.4-4), łączące dane z modeli CAD, wykorzystują podobne zautomatyzowane metody obliczania czasu, które znacznie skracają proces podejmowania decyzji. Pozwala to na błyskawiczne i dokładne planowanie harmonogramów prac i obliczanie całkowitego czasu wymaganego do wykonania wszystkich zadań w ramach realizacji projektu, z uwzględnieniem rzeczywistych cen.

Rozszerzone warstwy atrybutów 6D -8D: od efektywności energetycznej do zapewnienia bezpieczeństwa

6D, 7D i 8D to rozszerzone poziomy modelowania informacji, z których każdy wnosi dodatkowe warstwy atrybutów do kompleksowego modelu informacyjnego projektu, którego podstawą są atrybuty modelu 3D wraz z ich liczbą i zakresem. Każda dodatkowa warstwa wprowadza określone parametry, które są wymagane do dalszego grupowania lub dalszej identyfikacji w innych systemach, takich jak na przykład systemy zarządzania nieruchomościami (PMS), wspomagane komputerowo zarządzanie obiektami (CAFM), zarządzanie projektami budowlanymi (CPM) i systemy zarządzania bezpieczeństwem (SMS).



Rys. 5.3-4 Atrybuty 6D, 7D i 8D w modelu informacji o danych rozszerzają uwzględnienie różnych aspektów projektu, od efektywności energetycznej po bezpieczeństwo.

- W **6D** oprócz bazy danych projektu (lub ramki danych (Rys. 4.1-13)) z geometrycznymi i objętościowymi atrybutami elementów, dodawane są informacje (kolumny atrybutów) dotyczące zrównoważonego rozwoju środowiska. Obejmuje to informacje związane z efektywnością energetyczną, śladem węglowym, możliwością recyklingu materiałów i wykorzystaniem technologii przyjaznych dla środowiska. Dane te umożliwiają ocenę wpływu projektu na środowisko, optymalizację decyzji projektowych i osiągnięcie celów zrównoważonego rozwoju (ESG).
- Atrybuty **7D** uzupełniają atrybuty wymagane do zarządzania konserwacją budynku. Są to dane dotyczące harmonogramów konserwacji, cykli życia komponentów, dokumentacji technicznej i historii napraw. Ten zestaw informacji zapewnia, że model może być zintegrowany z systemami konserwacji (CAFM, AMS), umożliwia efektywne planowanie konserwacji, wymianę sprzętu i zapewnia wsparcie przez cały cykl życia obiektu.
- **8D** dodatkowa warstwa atrybutów, - zawiera informacje związane z bezpieczeństwem - zarówno na etapie budowy, jak i podczas późniejszej eksploatacji. Model obejmuje środki bezpieczeństwa personelu, instrukcje awaryjne, systemy ewakuacji i wymogi ochrony przeciwpożarowej. Integracja tych danych z modelem cyfrowym pomaga z wyprzedzeniem uwzględnić ryzyko i opracować rozwiązania architektoniczne, inżynierijne i organizacyjne, które uwzględniają wymogi bezpieczeństwa i higieny pracy.

W ustrukturyzowanej formie tabelarycznej warstwy od 4D do 8D reprezentują dodatkowe atrybuty w postaci kolumn z wypełnionymi wartościami (Rys. 5.3-5) dodanymi do już wypełnionych atrybutów modelu 3D, takich jak nazwa, kategoria, typ i charakterystyka objętościowa. Wartości w warstwach atrybutów 6D, 7D i 8D zawierają dodatkowe dane tekstowe i liczbowe, takie jak procent recyklingu, ślad węglowy, okres gwarancji, cykl wymiany, data instalacji, protokoły bezpieczeństwa itp.



ID	Type Name	Width	Length	Recyclability	Carbon Footprint	Warranty Period	Replacement Cycle	Maintenance Schedule	Installation Date	Wellbeing Factors	Safety Protocols
W-NEW	Window	120 cm	-	90%	1622 kgCO ₂ e	8 years	20 years	Annual	-mon	XYZ Windows	ISO 45001
W-OLD1	Window	100 cm	140 cm	90%	1522 kgCO ₂ e	8 years	15 years	Biannual	08/22/2024	XYZ Windows	OSHA Standard
W-OLD2	Window	110 cm	160 cm	90%	1522 kgCO ₂ e	-	15 years	Biannual	08/24/2024	????	OSHA Standard
D-122	Door	90 cm	210 cm	100%	1322 kgCO ₂ e	15 years	25 years	Biannual	08/25/2024	Doors Ltd.	OSHA Standard

Rys. 5.3-5 6D -8D dodaje warstwy atrybutów do modelu informacji o danych, który zawiera już atrybuty geometryczne i objętościowe z modelu 3D.

Dla naszego nowego okna (Rys. 4.4-1), element o identyfikatorze W-NEW (Rys. 5.3-5) może mieć następujące atrybuty 3D -8D:

3D - atrybuty - informacje geometryczne uzyskane z systemów CAD:

- "Nazwa typu" - element "Okno"
- "Szerokość" - 120 cm
- Dodatkowo można dodać punkty "Bounding Box" elementu lub jego "geometrii BREP / MESH" jako osobny atrybut.

Atrybuty 6D - zrównoważony rozwój środowiskowy:

- Wskaźnik recyklingu na poziomie 90 procent
- "Ślad węglowy" - 1 622 kg CO₂

Atrybuty 7D - dane dotyczące zarządzania obiektem:

- "Okres gwarancji" - 8 lat
- "Cykl wymiany" ma już 20 lat
- "Konserwacja" - wymagana corocznie

Atrybuty 8D - zapewnienie bezpiecznego użytkowania i eksploatacji budynków:

- Okno "Zainstalowano" - przez firmę "XYZ Windows"
- "Standard bezpieczeństwa" - zgodny z normą ISO 45001

Wszystkie parametry zapisane w bazie danych lub zbiorze danych (rys. 5.3-5) są potrzebne specjalistom z różnych działów do grupowania, wyszukiwania lub obliczeń. Taki wielowymiarowy, oparty na atrybutach opis obiektów projektu zapewnia pełny obraz ich cyklu życia, wymagań operacyjnych i wielu innych aspektów niezbędnych do projektowania, budowy i eksploatacji projektu.

Szacowanie emisji CO₂ i obliczanie emisji dwutlenku węgla z projektów budowlanych

Oprócz tematu zrównoważonego rozwoju projektów budowlanych na etapie 6D (Rys. 5.3-5), nowoczesne budownictwo koncentruje się na zrównoważonym rozwoju środowiskowym projektów, gdzie jednym z kluczowych aspektów staje się ocena i minimalizacja emisji dwutlenku węgla CO₂, które występują na etapach cyklu życia projektu (np. produkcja i instalacja).

Szacowanie i obliczanie emisji dwutlenku węgla z materiałów budowlanych jest procesem, w którym całkowita emisja dwutlenku węgla jest określana poprzez pomnożenie objętościowych atrybutów elementu lub grupy elementów wykorzystywanych w projekcie przez odpowiedni współczynnik emisji dwutlenku węgla dla danej kategorii.

Uwzględnienie emisji dwutlenku węgla podczas oceny projektów budowlanych w ramach szerszych kryteriów ESG (środowiskowych, społecznych i ładu korporacyjnego) dodaje nowy poziom złożoności do analizy. Jest to szczególnie ważne dla klienta-inwestora przy uzyskiwaniu odpowiednich certyfikatów, takich jak LEED® (Leadership in Energy and Environmental Design), BREEAM® (Building Research Establishment Environmental Assessment Method) lub DGNB® (Deutsche Gesellschaft für Nachhaltiges Bauen). Uzyskanie jednego z tych certyfikatów może znacznie zwiększyć atrakcyjność rynkową nieruchomości, uprościć oddanie do użytku i zapewnić zgodność z wymogami najemców zorientowanych na zrównoważony rozwój (ESG). W zależności od wymagań projektu, można również zastosować HQE (Haute Qualité Environnementale, francuski standard zielonego budynku), WELL (WELL Building Standard, koncentrujący się na zdrowiu i komforcie użytkowników) oraz GRESB (Global Real Estate Sustainability Benchmark)

Environmental, social and governance **ESG** (environmental-mental, social and governance) to szeroki zestaw zasad, które można wykorzystać do oceny ładu korporacyjnego, wpływu społecznego i środowiskowego firmy zarówno wewnętrznie, jak i zewnętrznie.

ESG, pierwotnie opracowany na początku XXI wieku przez fundusze finansowe w celu dostarczenia inwestorom informacji na temat szerokich kryteriów środowiskowych, społecznych i ładu korporacyjnego, stał się kluczowym wskaźnikiem oceny zarówno spółek, jak i projektów, w tym projektów budowlanych. Według badań przeprowadzonych przez główne firmy konsultingowe, kwestie środowiskowe, społeczne i związane z ładem korporacyjnym (ESG) stają się integralną częścią branży budowlanej.

Według EY (2023) "The Path to Carbon Neutrality", firmy, które aktywnie wdrażają zasady ESG, nie tylko zmniejszają długoterminowe ryzyko, ale także zwiększają efektywność swoich modeli biznesowych, co jest szczególnie ważne w globalnej transformacji rynków [103]. W raporcie ESG Awareness firmy PwC zauważono, że świadomość firm w zakresie znaczenia czynników ESG wahala się od 67% do 97%, przy czym większość organizacji postrzega te trendy jako kluczowe dla przyszłego zrównoważonego rozwoju [104], a firmy w większości dostrzegają znaczną presję ze strony interesariuszy na integrację zasad ESG.

Tak więc integracja zasad ESG z projektami budowlanymi nie tylko przyczynia się do uzyskania międzynarodowych certyfikatów zrównoważonego rozwoju, takich jak LEED, BREEAM, DGNB, ale także zapewnia długoterminowy zrównoważony rozwój i konkurencyjność firm w branży.

Jednym z najważniejszych czynników wpływających na ogólny ślad węglowy projektu budowlanego są etapy produkcji i logistyki materiałów i komponentów budowlanych. Materiały wykorzystywane na placu budowy często mają decydujący wpływ na całkowitą emisję CO₂, zwłaszcza na wczesnych etapach cyklu życia projektu - od wydobycia surowców po dostawę na plac budowy.

Obliczanie emisji według kategorii lub typu elementu budynku wymaga zastosowania referencyjnych wskaźników emisji dwutlenku węgla, które odzwierciedlają ilość CO₂ generowanego przy produkcji różnych materiałów. Materiały te obejmują beton, cegły, stal z recyklingu, aluminium i inne. Wartości te są zazwyczaj pobierane z renomowanych źródeł i międzynarodowych baz danych, takich jak UK ICE 2015 (Inventory of Carbon and Energy) i US EPA 2006 (U.S. Environmental Protection Agency) [105]. Poniższa tabela (Rys. 5.3-6) podsumowuje bazowe wskaźniki emisji dla szeregu popularnych materiałów budowlanych. Dla każdego materiału podano dwa kluczowe parametry: konkretne emisje CO₂ (w kilogramach na kilogram materiału) oraz współczynniki konwersji objętości na masę (w kilogramach na metr sześcienny), które są niezbędne do zintegrowania obliczeń z modelem projektowym i powiązania z grupowaniem danych QTO.



Carbon Emitted in Production		UK ICE Database (2015) USEPA (2006)	UK ICE Database (2015) USEPA (2006)	Coefficient m³ to kg
Material	Abbreviated	Process Emissions (kg CO ₂ / kg of product) (K1)	Process Emissions (kg CO ₂ / kg of product) (K2)	
Concrete	Concrete	0.12	0.12	2400
Concrete block	Concrete_block	0.13**	0.14	2000
Brick	Brick	0.24	0.32	2000
Medium density fiberboard (MDF)	MDF	0.39*	0.32	700
Recycled steel (avg recy content)	Recycled_steel	0.47	0.81	7850
Glass (not including primary mfg.)	Glass	0.59	0.6	2500
Cement (Portland, masonry)	Cement	0.95	0.97	1440
Aluminum (virgin)	Aluminum	12.79	16.6	2700

Rys. 5.3-6 Ilość dwutlenku węgla emitowanego podczas produkcji różnych materiałów budowlanych, zgodnie z bazą danych UK ICE i US EPA.

Aby obliczyć całkowitą emisję CO₂ dla projektu, podobnie jak w przypadku obliczeń 4D i 5D, konieczne jest określenie objętości atrybutów każdej grupy obiektów. Można to zrobić za pomocą narzędzi analizy ilościowej (QTO), uzyskując objętości atrybutów w metrach sześciennych, jak omówiono szczegółowo w sekcji dotyczącej Quantity take-off. Objętości te są następnie mnożone przez odpowiednie współczynniki dla atrybutu "emisje CO₂ z procesów" dla każdej grupy materiałów.

- » Automatycznie wyodrębnijmy tabelę objętości według typu elementu z projektu CAD (BIM), grupując wszystkie dane projektu, jak to już zrobiono w poprzednich rozdziałach. Aby wykonać to zadanie, odwołajmy się do LLM.

Prosimy o pogrupowanie tabeli DataFrame z projektu CAD (BIM) według parametru kolumny "Nazwa obiektu" (lub "Typ") i pokazanie liczby elementów w każdej grupie oraz podsumowanie parametru "Objętość" dla wszystkich elementów w typie. ↴

▣ Odpowiedź LLM :

```

Create in LLM Chat
Data
f(x)
Output
Run in IDE
  
```

```

1 import pandas as pd
2
3 # Import Projekt as CSV twin DataFrame
4 df = pd.read_csv('your_dataframe.csv')
5
6 # Group, count, and sum 'Volume' in one line
7 result = df.groupby('ObjectType')['Volume'].agg({'Amount': 'size', 'Sum of Volumes': 'sum'})
  
```

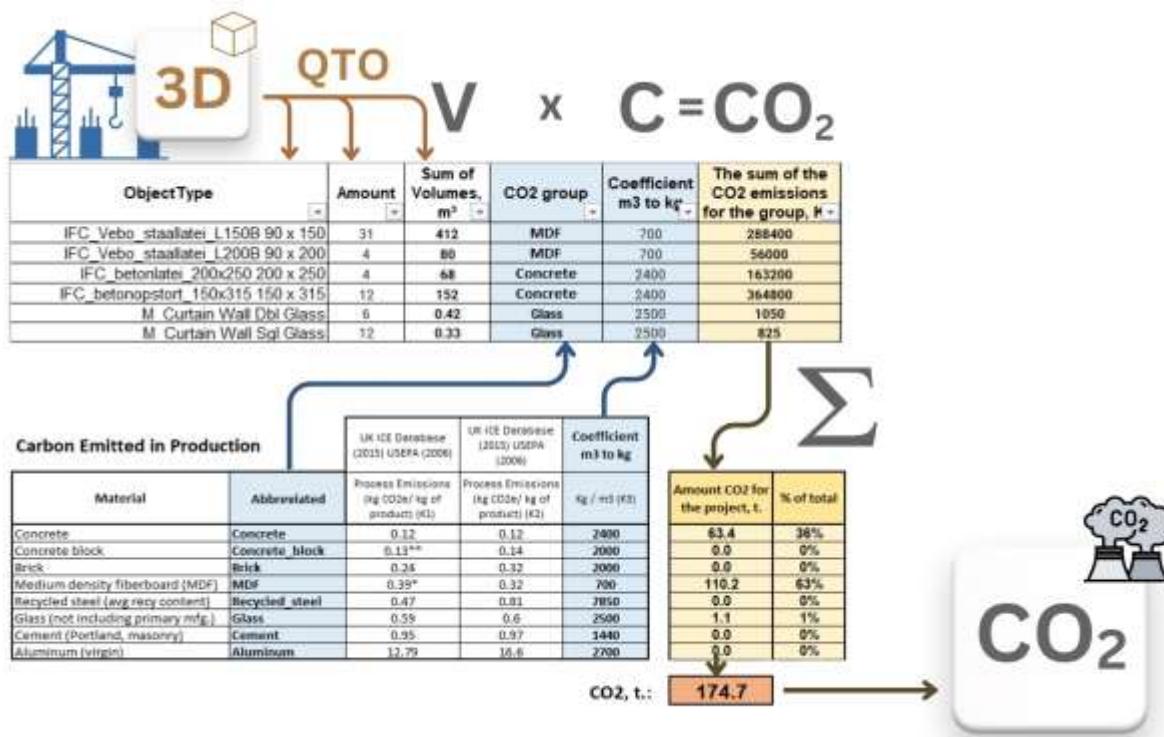
ObjectType	Amount	Sum of Volumes, m ³
IFC_Vebo_staalplatei_L150B_90 x 150	31	412
IFC_Vebo_staalplatei_L200B_90 x 200	4	80
IFC_betonplatei_200x250_200 x 250	4	68
IFC_betonopstort_150x315_150 x 315	12	152
M_Curtain Wall Dbl Glass	6	0.42
M_Curtain Wall Sgl Glass	12	0.33

Rys. 5.3-7 Wygenerowany kod w LLM pogrupował dla nas encje projektu według typu (ObjectType) z podsumowanym atrybutem "Volume".

Aby zautomatyzować obliczanie całkowitej emisji CO₂ dla całego projektu, wystarczy skonfigurować automatyczne mapowanie danych w tabeli lub ręcznie powiązać typy elementów (Rys. 5.3-7) z odpowiednimi typami materiałów (Rys. 5.3-6) z tabeli współczynników emisji. Gotową tabelę ze współczynnikami emisji i wzorami, a także kod do pobierania objętości z formatów CAD (BIM) i automatyzacji określania CO₂ można znaleźć w serwisie GitHub, wyszukując "CO₂_calculating-the-embodyied-carbon. DataDrivenConstruction." [106].

Tak więc integracja danych po zgrupowaniu elementów QTO z bazy danych CAD umożliwia automatyczne obliczanie emisji dwutlenku węgla (rys. 5.3-8) dla różnych opcji projektowych. Umożliwia to analizę wpływu różnych materiałów w różnych wariantach i wybór tych rozwiązań, które spełniają wymagania dotyczące emisji CO₂, w celu uzyskania określonego certyfikatu po oddaniu budynku do użytku.

Szacowanie emisji CO₂ poprzez mnożenie współczynników przez objętości zgrupowanych elementów projektu jest typowym przykładem zadania w procesie uzyskiwania przez firmę budowlaną oceny ESG (np. certyfikacji LEED) dla obiektu.



Rys. 5.3-8 Integracja grup QTO z baz danych CAD zapewnia dokładność i automatyzację w uzyskiwaniu szacunków ostatecznych emisji CO₂.

Podobnie, definiując objętości grup elementów, możemy wykonywać obliczenia dla kontroli materiałów i logistyki, monitorowania i zarządzania jakością, modelowania i analizy energii oraz wielu innych zadań w celu uzyskania nowego statusu atrybutu (parametru w tabeli) zarówno dla poszczególnych grup elementów, jak i całego projektu.

Jeśli liczba takich procesów obliczeniowych w firmie zaczyna rosnąć, pojawia się pytanie o potrzebę automatyzacji takich obliczeń i wdrożenia wyników obliczeń do procesów firmy i systemów zarządzania danymi.

Ze względu na złożoność kompletnego rozwiązania, średnie i duże firmy z branży budowlanej zlecają taką automatyzację firmom zajmującym się tworzeniem systemów ERP (lub PMIS). Firmy deweloperskie tworzą jeden kompleksowy system modułowy dla dużych klientów do zarządzania wieloma różnymi warstwami informacji, w tym obliczeniami materiałów i zasobów.



ROZDZIAŁ 4.4.

SYSTEMY BUDOWLANE ERP I PMIS

Construction ERP -systemy na przykładzie kalkulacji i kosztorysów

Modułowe systemy ERP integrują różne warstwy atrybutów (informacji) i przepływy danych w jeden kompleksowy system, umożliwiając kierownikom projektów zarządzanie zasobami, finansami, logistiką i innymi aspektami projektu w sposób zsynchronizowany w ramach jednej platformy. Budowlany system ERP działa jak "mózg" projektów budowlanych, upraszczając powtarzalne procesy poprzez automatyzację, zapewniając przejrzystość i kontrolę w całym procesie budowlanym.

Systemy ERP dla budownictwa (Enterprise Resource Planning) to kompleksowe rozwiązania programowe zaprojektowane do zarządzania i optymalizacji różnych aspektów procesu budowlanego. Rdzeniem systemów ERP dla budownictwa są moduły do zarządzania kosztorysami i harmonogramami, co czyni je ważnym narzędziem do efektywnego planowania zasobów.

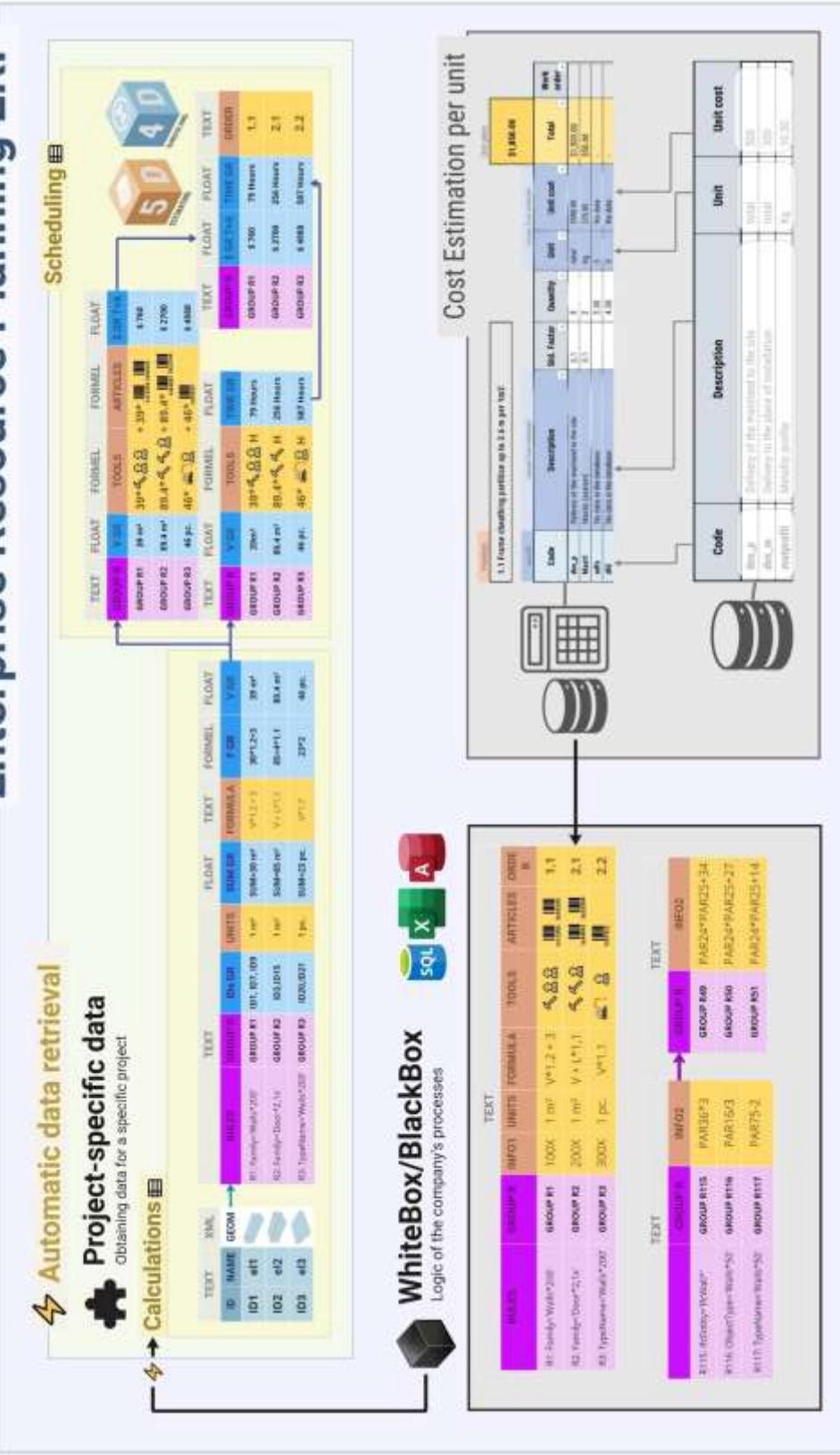
Moduły ERP -systemy umożliwiają użytkownikom wprowadzanie, przetwarzanie i analizowanie danych w ustrukturyzowany sposób obejmujący różne aspekty projektu, które mogą obejmować rozliczanie kosztów materiałów i robocizny, wykorzystanie sprzętu, zarządzanie logistiką, zasoby ludzkie, kontakty i inne działania budowlane.

Jednym z bloków funkcjonalnych systemu jest moduł automatyzacji logiki biznesowej - BlackBox/WhiteBox, który pełni rolę centrum sterowania procesem.

BlackBox /WhiteBox umożliwia specjalistom korzystającym z systemu ERP elastyczne zarządzanie, za pomocą praw dostępu, różnymi aspektami działalności, które zostały już wcześniej skonfigurowane przez innych użytkowników lub administratorów. W kontekście systemów ERP terminy *BlackBox* i *WhiteBox* odnoszą się do poziomów przejrzystości i możliwości kontroli wewnętrznej logiki systemu:

- **BlackBox** ("czarna skrzynka") - użytkownik wchodzi w interakcję z systemem za pośrednictwem interfejsu, bez dostępu do wewnętrznej logiki wykonywania procesu. System wykonuje obliczenia samodzielnie, w oparciu o predefiniowane reguły ukryte przed użytkownikiem końcowym. Użytkownik wprowadza dane i otrzymuje wynik, nie wiedząc, jakie atrybuty lub współczynniki zostały użyte wewnętrznie.
- **WhiteBox** ("biała skrzynka") - logika procesu jest dostępna do przeglądania, dostosowywania i modyfikacji. Zaawansowani użytkownicy, administratorzy lub integratorzy mogą ręcznie definiować algorytmy przetwarzania danych, reguły obliczeniowe i scenariusze interakcji między podmiotami projektu.

Enterprise Resource Planning ERP



Rys. 5.4-1 Architektura budowlanego systemu ERP -system, aby uzyskać szacunki i harmonogramy prac poprzez ręczne wypełnianie atrybutów objętości.

Przykładem może być sytuacja, w której doświadczony użytkownik lub administrator ustala reguły: które atrybuty w kosztorysie powinny zostać pomnożone między sobą lub pogrupowane według określonej cechy i gdzie powinien zostać zapisany wynik końcowy. Następnie mniej wyszkoleni specjalisci, tacy jak inżynierowie kosztorysujący, po prostu ładują nowe dane do ERP za pośrednictwem interfejsu użytkownika - i otrzymują gotowe szacunki, harmonogramy lub specyfikacje bez konieczności pisania kodu lub rozumienia szczegółów technicznych logiki.

W poprzednich rozdziałach moduły obliczeniowe i logiczne zostały omówione w kontekście interakcji LLM. W środowisku ERP takie obliczenia i transformacje odbywają się w ramach modułów ukrytych za interfejsem przycisków i formularzy.

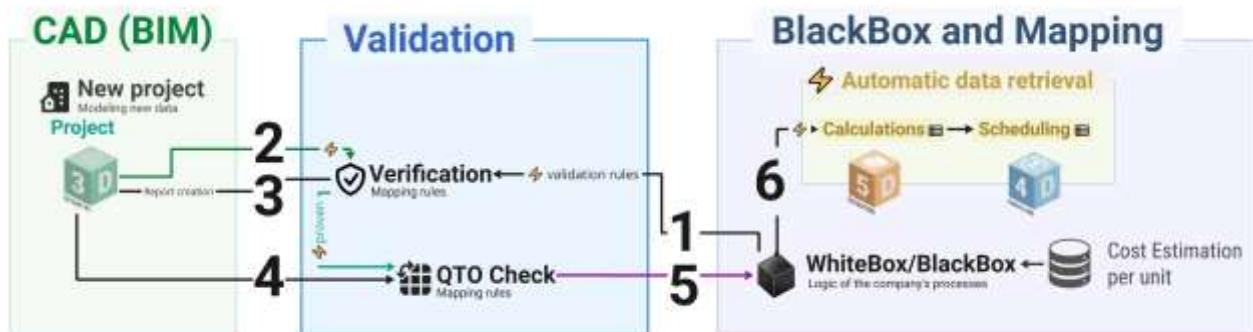
W poniższym przykładzie (rys. 5.4-1) administrator systemu ERP w module BlackBox /WhiteBox zdefiniował reguły dopasowywania atrybutów podmiotów z szacunków do atrybutów grupowania QTO. Dzięki tak skonfigurowanemu (przez kierownika lub administratora) modułowi BlackBox/WhiteBox użytkownik (kosztorysant lub inżynier), dodając ręcznie atrybut ilości lub objętości za pośrednictwem interfejsu użytkownika ERP, automatycznie otrzymuje gotowe kosztorysy i harmonogramy prac. W ten sposób procesy obliczania i generowania kosztorysów omówione w poprzednich rozdziałach za pomocą kodu wewnętrz systemu ERP stają się półautomatycznym przenośnikiem.

Połączenie tego półautomatycznego procesu z atrybutami objętościowymi z modeli CAD (BIM) (rys. 4.1-13), na przykład poprzez załadowanie projektu CAD do modułu ERP, który jest wstępnie skonfigurowany do tego celu, zamienia przepływ danych w zsynchronizowany mechanizm zdolny do autonomicznej i natychmiastowej aktualizacji wartości poszczególnych grup elementów lub całego projektu w odpowiedzi na wszelkie zmiany w nim w fazie projektowania, podczas ładowania modelu CAD do ERP.

W celu stworzenia zautomatyzowanego przepływu danych (rys. 5.4-2) pomiędzy systemami CAD (BIM) i ERP, podstawowe procesy i wymagania dla danych z baz danych modeli CAD (BIM) muszą być zdefiniowane w ustrukturyzowany sposób, jak już omówiliśmy w powyższym rozdziale "Wymagania i zapewnienie jakości danych". Proces ten w ERP jest podzielony na podobne kroki:

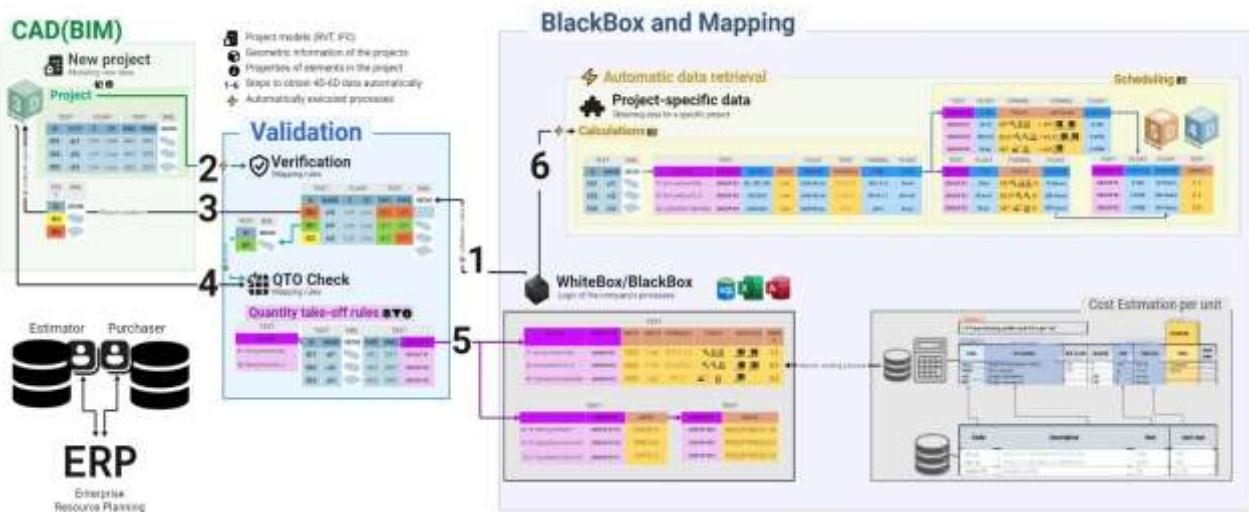
- **Tworzenie reguł walidacji (1)**, które odgrywają ważną rolę w zapewnianiu dokładności danych wprowadzanych do systemu ERP. Reguły walidacji służą jako filtry, które sprawdzają poprawność jednostek i ich atrybutów, umożliwiając wejście do systemu tylko tym elementom, które spełniają wymagania. Więcej informacji na temat weryfikacji i walidacji można znaleźć w rozdziale "Tworzenie wymagań i walidacja jakości danych".
- Następnie w systemie ERP **odbywa się proces weryfikacji (2)**, który potwierdza, że wszystkie elementy podmiotu projektu wraz z ich atrybutami i wartościami zostały utworzone poprawnie i są gotowe do kolejnych etapów przetwarzania.
- Jeśli wystąpią problemy z niekompletnymi danymi atrybutów, **generowany jest raport (3)**, a projekt, wraz z instrukcjami dotyczącymi korekty, jest wysyłany do korekty, aż będzie gotowy do następnej iteracji.
- Po zatwierdzeniu i zweryfikowaniu danych projektu są one wykorzystywane w innym module ERP **(4) do tworzenia tabel Quantity Take-Off (QTO)**, które tworzą atrybuty ilości dla grup podmiotów, materiałów i zasobów zgodnie z wcześniej wygenerowanymi regułami (WhiteBox/BlackBox).
- Dane pogrupowane według reguł dopasowywania lub QTO są automatycznie **integrowane z obliczeniami (np. kosztów i czasu) (5)**.

- W ostatnim kroku systemu ERP użytkownik, mnożąc atrybuty zakresu z tabeli QTO z atrybutami tabel procesów (np. szacowane pozycje), **automatycznie generuje wyniki obliczeń (6)** (np. szacunki kosztów, harmonogramy prac lub emisje CO₂) dla każdej grupy podmiotów i dla projektu jako całości.

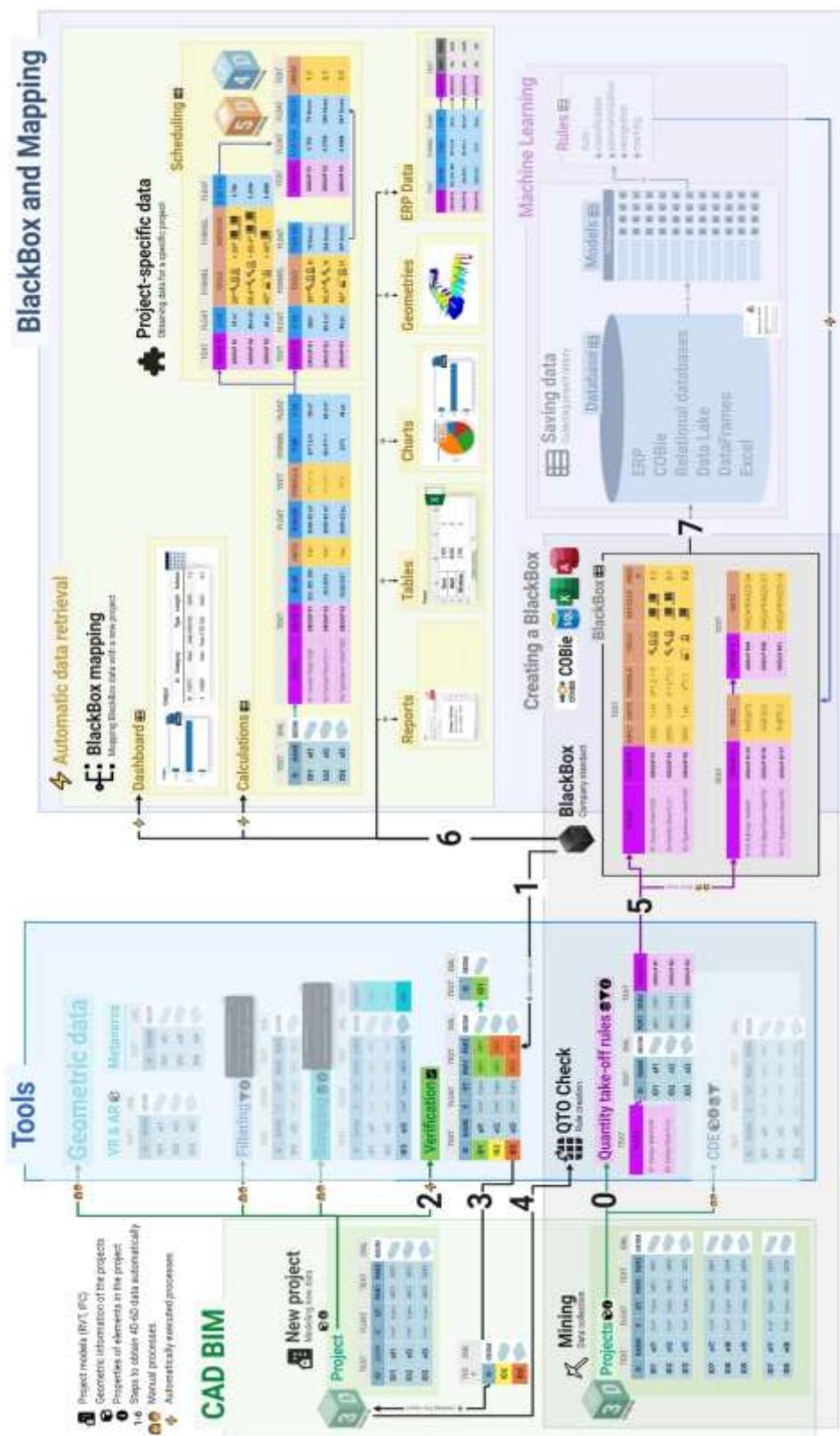


Rys. 5.4-2 Architektura budowlanego systemu ERP z CAD (BIM), od tworzenia reguł walidacji (1) do automatycznego obliczania kosztów i harmonogramów prac (5-6).

W modułowym systemie ERP procesy są zintegrowane za pomocą oprogramowania zawierającego interfejs użytkownika. Za interfejsem znajduje się zaplecze, w którym ustrukturyzowane tabele przetwarzają dane, wykonując różne operacje, które zostały wstępnie skonfigurowane przez menedżera lub administratora. W rezultacie użytkownik, dzięki wstępnie zdefiniowanej i dostosowanej logice automatyzacji (w modułach BlackBox /WhiteBox), otrzymuje półautomatycznie przygotowane dokumenty, które spełniają jego zadania



Rys. 5.4-3 ERP -system pomaga menedżerom i użytkownikom poruszać się między specjalistycznymi tabelami w celu generowania nowych danych.



Rys. 5.4-4 ERP -system jest zintegrowany z narzędziami analitycznymi i automatyzuje proces podejmowania decyzji w firmie.

Podobnie, procesy w systemach ERP, od etapu początkowego do końcowych obliczeń (kroki 1-6, Rys. 5.4-3) są łańcuchem powiązanych ze sobą kroków, które ostatecznie zapewniają przejrzystość, wydajność i dokładność planowania.

Nowoczesne systemy ERP dla budownictwa obejmują nie tylko moduły kalkulacji kosztów i harmonogramów, ale także dziesiątki innych wstępnie skonfigurowanych modułów, zwykle obejmujących zarządzanie dokumentami, śledzenie postępów projektu, zarządzanie umowami, łańcuch dostaw i logistyką, a także integrację z innymi systemami i platformami biznesowymi. Zintegrowane narzędzia analityczne ERP pozwalają użytkownikom zautomatyzować tworzenie pulpitów nawigacyjnych do monitorowania kluczowych wskaźników wydajności projektu (KPI - key performance indicators). Zapewnia to scentralizowane i spójne zarządzanie wszystkimi aspektami projektu budowlanego, z próbą połączenia dużej liczby aplikacji i systemów na jednej platformie.

W przyszłości systemy ERP-analytics będą wykorzystywane w połączeniu z uczeniem maszynowym w celu zwiększenia dokładności i optymalizacji procesu obliczania przyszłych atrybutów projektu. Dane i atrybuty analizowane i gromadzone z systemów ERP w Big Data (rys. 5.4-4) będą w przyszłości stanowić podstawę do tworzenia modeli predyktacyjnych, które mogą dokładnie przewidywać potencjalne opóźnienia, zagrożenia lub, na przykład, możliwe zmiany w kosztach materiałów.

Jako alternatywę dla ERP, branża budowlana często wykorzystuje PMIS (Project Management Information System), system zarządzania projektami przeznaczony do szczegółowej kontroli zadań na poziomie indywidualnego projektu budowlanego.

PMIS: Pośrednik między ERP a placem budowy

W przeciwieństwie do ERP, który obejmuje cały łańcuch procesów biznesowych firmy, PMIS koncentruje się na zarządzaniu konkretnym projektem, monitorowaniu harmonogramów, budżetów, zasobów i dokumentacji.

PMIS (Project Management Information System) to oprogramowanie do zarządzania projektami budowlanymi przeznaczone do planowania, śledzenia, analizowania i raportowania wszystkich aspektów projektu.

PMIS umożliwia zarządzanie dokumentami, harmonogramami, budżetami i na pierwszy rzut oka PMIS może wydawać się duplikatem rozwiązania ERP, ale kluczową różnicą jest poziom zarządzania:

- **ERP** koncentruje się na procesach biznesowych firmy jako całości: kosztach, kontraktach, zaopatrzeniu, zasobach ludzkich i zarządzaniu zasobami na poziomie korporacyjnym.
- **PMIS** koncentruje się na zarządzaniu indywidualnymi projektami, zapewniając szczegółowe planowanie, kontrolę zmian, raportowanie i koordynację uczestników.

W wielu przypadkach to systemy ERP mają już wystarczającą funkcjonalność, a wdrożenie PMIS staje się bardziej kwestią wygody i preferencji firmy. Wielu wykonawców i klientów korzysta z PMIS nie dlatego, że jest to konieczne, ale dlatego, że jest to narzucone przez dostawcę lub dużego klienta, który chce agregować dane na określonej platformie.

Należy wspomnieć, że w międzynarodowej terminologii zarządzania projektami budowlanymi istnieją inne odrębne popularne pojęcia, takie jak PLM (Product Lifecycle Management) oraz EPC i EPC-M (Engineering,

Procurement and Construction Management) - metody kontraktowania w branży budowlanej.

Jeśli firma korzysta już z ERP z modułami zarządzania projektami, wprowadzenie PMIS może być niepotrzebnym łącznikiem, który powiela funkcjonalność. Jeśli jednak procesy nie są zautomatyzowane, a dane są rozproszone, PMIS może być wygodniejszym i łatwiejszym w utrzymaniu narzędziem.

Spekulacja, zysk, wyspiarski charakter i brak przejrzystości w ERP i PMIS

Pomimo zewnętrznej prostoty interfejsów i procedur, budowlane systemy ERP i PMIS są w większości przypadków zamkniętymi i nieelastycznymi rozwiązaniami. Takie systemy są zwykle dostarczane jako wstępnie skonfigurowany pakiet oprogramowania od jednego dostawcy, z ograniczonym dostępem do wewnętrznych baz danych i logiki procesów.

Dostawcy systemów CAD (BIM) oraz częściej przejmują rozwój i kontrolę nad takimi systemami, ponieważ ich bazy danych zawierają informacje wymagane przez systemy ERP: ilościowe i objętościowe atrybuty elementów projektu. Jednak zamiast zapewniać dostęp do tych danych w formacie otwartym lub nadającym się do odczytu maszynowego, dostawcy oferują jedynie ograniczone scenariusze użytkownika i zamkniętą logikę przetwarzania - predefiniowaną w modułach BlackBox. Zmniejsza to elastyczność systemu i uniemożliwia jego dostosowanie do konkretnych warunków projektu.

Ograniczona przejrzystość danych pozostaje jednym z kluczowych wyzwań cyfrowych procesów w budownictwie. Zamknięta architektura baz danych, brak dostępu do pełnych zestawów atrybutów elementów budynku, skupienie się na modułach automatyzacji BlackBox i brak otwartych interfejsów znacznie zwiększa ryzyko biurokracji dokumentowej. Takie ograniczenia tworzą wąskie gardła w procesie decyzyjnym, utrudniają weryfikację informacji i otwierają drzwi do ukrywania danych lub spekulacji w systemach ERP/PMIS. Użytkownicy zazwyczaj otrzymują jedynie ograniczony dostęp - czy to w postaci określonego interfejsu, czy częściowego API - bez możliwości bezpośredniej interakcji z podstawowymi źródłami danych. Jest to szczególnie istotne w przypadku parametrów generowanych automatycznie z projektów CAD, takich jak objętości, powierzchnie i ilości wykorzystywane do obliczeń QTO.

W konsekwencji, zamiast poszukiwać efektywności poprzez automatyzację procesów, otwarte dane, redukcję kosztów transakcyjnych i tworzenie nowych modeli biznesowych, wiele firm budowlanych koncentruje się na zarządzaniu parametrami zewnętrznymi - manipulując czynnikami, współczynnikami korygującymi i metodami kalkulacji wpływającymi na koszty projektu w zamkniętych platformach ERP/PMIS. Stwarza to pole do spekulacji, zniekształca rzeczywiste koszty produkcji i zmniejsza zaufanie między wszystkimi uczestnikami procesu budowlanego.

W budownictwie zysk powstaje jako różnica między przychodami z zakończonego projektu a kosztami zmiennymi, które obejmują projekt, materiały, robociznę i inne koszty bezpośrednio bezpośrednio związane z realizacją projektu. Kluczowym czynnikiem wpływającym na wartość tych kosztów jest jednak nie tylko technologia czy logistyka, ale także szybkość i dokładność obliczeń oraz jakość decyzji zarządczych w firmie.

Problem ten pogłębia fakt, że w większości firm budowlanych procesy kalkulacji kosztów pozostają nieprzejrzyste nie tylko dla klientów, ale także dla samych pracowników, którzy nie są częścią działów kosztorysowania lub finansów. Taka zamkniętość sprzyja tworzeniu się w firmie uprzywilejowanej grupy specjalistów - nosicieli "wiedzy finansowej", którzy mają wyłączne prawo do edycji atrybutów i współczynników korygujących w systemach ERP/PMIS. Pracownicy ci, wraz z szefami firm, mogą faktycznie kontrolować logikę finansową projektu.

W takich warunkach kosztorysanci zamieniają się w "finansowych żonglerów", balansujących między maksymalizacją zysku firmy a potrzebą utrzymania konkurencyjnej ceny dla klienta. Jednocześnie muszą unikać rażących i rażących manipulacji, aby nie podważać reputacji firmy. To właśnie na tym etapie ustalone są współczynniki ukrywające zawyżone ilości lub koszty materiałów i prac.

W rezultacie głównym schematem zwiększania efektywności i rentowności firm działających w branży budowlanej nie jest automatyzacja i przyspieszenie procesów decyzyjnych, ale spekulacja cenami materiałów i robót (rys. 5.4-5). Zawyżanie kosztów robót i materiałów odbywa się poprzez szarą księgowość w zamkniętych systemach ERP /PMIS - poprzez zawyżanie procentów w stosunku do średnich cen rynkowych materiałów lub ilości robót za pomocą współczynników (rys. 5.1-6), które zostały omówione w rozdziale "Kompilacja kalkulacji i kalkulacja kosztów robót na podstawie bazy zasobów".

W efekcie klient otrzymuje kalkulację, która nie odzwierciedla rzeczywistych kosztów czy zakresu prac, a jest pochodną wielu ukrytych wewnętrznych współczynników. Jednocześnie podwykonawcy, próbując sprostać zaniżonym stawkom ustalonym przez generalnego wykonawcę, często zmuszeni są do zakupu tańszych i niskiej jakości materiałów, co pogarsza ostateczną jakość budowy.

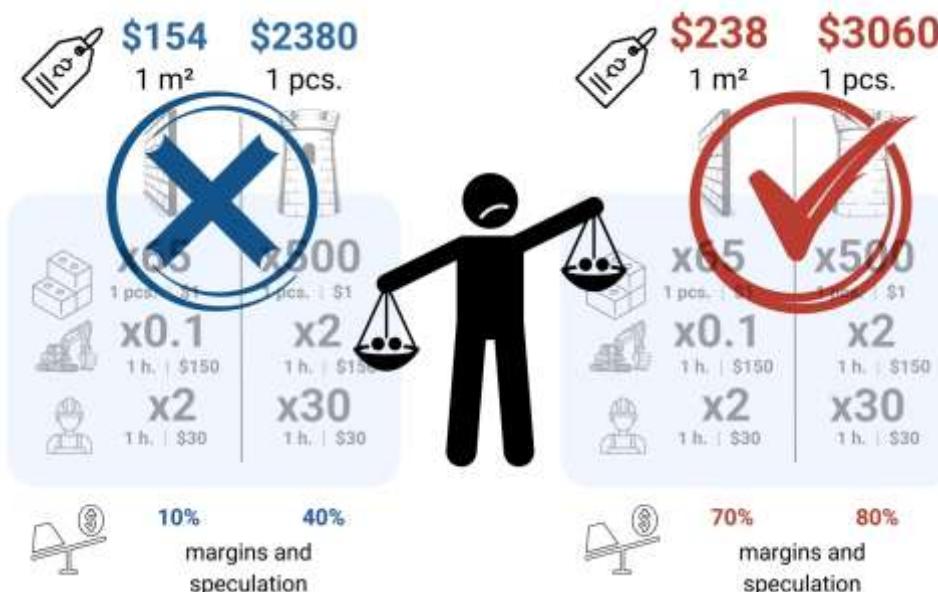
Spekulacyjny proces poszukiwania zysków z powietrza kończy się krzywdą zarówno dla klientów, którzy otrzymują niewiarygodne dane, jak i dla kadry kierowniczej, która jest zmuszona do poszukiwania coraz większej liczby modeli spekulacyjnych

W rezultacie, im większy projekt, tym wyższy poziom biurokracji w zarządzaniu danymi i procesami. Każdy krok i każdy moduł często kryje w sobie nieprzejrzyste współczynniki i narzuty wbudowane w algorytmy obliczeniowe i wewnętrzne procedury. To nie tylko utrudnia audyt, ale także znaczowo zniekształca obraz finansowy projektu. W dużych projektach budowlanych takie praktyki często prowadzą do wielokrotnego (czasami nawet dziesięciokrotnego) wzrostu kosztów końcowych, podczas gdy rzeczywiste wielkości i koszty pozostają poza skuteczną kontrolą klienta (rys. 2.1-3 Porównanie planowanych i rzeczywistych kosztów dużych projektów infrastrukturalnych w Niemczech).

Według raportu McKinsey & Company *Imagining the Digital Future of Construction* (2016), duże projekty budowlane są realizowane średnio 20% później niż planowano i do 80% powyżej budżetu [107].

Działalność kosztorysowania i budżetowania stają się najbardziej strzeżonym ogniwem w firmie. Dostęp do nich jest ścisłe ograniczony nawet dla wewnętrznych specjalistów, a ze względu na zamkniętą logikę i struktury baz danych niemożliwa jest obiektywna ocena efektywności decyzji projektowych bez zniekształceń. Brak

przejrzystości prowadzi do tego, że firmy nie są zmuszone do optymalizacji procesów, ale do walki o przetrwanie poprzez "kreatywne" zarządzanie liczbami i czynnikami (rys. 5.3-1, rys. 5.1-6 - na przykład parametr "Bid. Factor").



Rys. 5.4-5 Współczynniki spekulacji na poziomie rozliczeń to główny zysk firm i sztuka żonglowania między jakością pracy a reputacją.

Wszystko to poddaje w wątpliwość dalsze stosowanie zamkniętych systemów ERP/PMIS w budownictwie. W kontekście transformacji cyfrowej i rosnących wymagań dotyczących przejrzystości ze strony klientów (rys. 10.2-3), jest mało prawdopodobne, aby realizacja projektu w dłuższej perspektywie pozostała zależna od zastrzeżonych rozwiązań, które ograniczają elastyczność, utrudniają integrację i hamują rozwój biznesu.

I bez względu na to, jak korzystna dla firm budowlanych może być praca z silosami danych i nieprzejrzystymi danymi w zamkniętych bazach danych - nieuchronnie przyszłość branży budowlanej będzie wiązała się z przejściem na otwarte platformy, czytelne dla maszyn i przejrzyste struktury danych oraz automatyzację opartą na zaufaniu. Transformacja ta będzie napędzana z góry - pod presją klientów, organów regulacyjnych i społeczeństwa, coraz bardziej wymagających odpowiedzialności, zrównoważonego rozwoju, przejrzystości i wykonalności ekonomicznej.

Koniec ery zamkniętych systemów ERP /PMIS: branża budowlana potrzebuje nowego podejścia

Korzystanie z nieporęcznych modułowych systemów ERP/PMIS składających się z dziesiątek milionów linii kodu sprawia, że wszelkie zmiany w nich są niezwykle trudne. W takim przypadku przejście na nową platformę w obecności modułów już wcześniej skonfigurowanych dla firmy, dziesiątek tysięcy artykułów w bazach danych zasobów (rys. 5.1-3) i tysięcy gotowych obliczeń (rys. 5.1-6) zamienia się w kosztowny i długotrwały proces. Im więcej kodu i starszej architektury - tym wyższy poziom wewnętrznej nieefektywności, a każdy nowy projekt tylko pogorszy sytuację. W wielu firmach migracja danych i integracja nowych rozwiązań stają się wieloletnimi epopejami, którym towarzyszą ciągłe przeróbki i

niekończące się poszukiwania kompromisów. Rezultatem jest często powrót do starych, dobrze znanych platform, pomimo ich ograniczeń.

Jak podkreślono w niemieckim raporcie Black Book [108] na temat systemowych błędów w zarządzaniu danymi budowlanymi, fragmentacja informacji i brak scentralizowanego podejścia do zarządzania nimi jest kluczową przyczyną nieefektywności. Bez standaryzacji i integracji dane tracą swoją wartość, stając się raczej archiwum niż narzędziem zarządzania.

Główną przyczyną utraty jakości danych jest nieodpowiednie planowanie i kontrola projektów budowlanych, co często prowadzi do znacznego wzrostu kosztów. Sekcja Black Book "Focus: The Cost Explosion" analizuje kluczowe czynniki przyczyniające się do tych niepożądanych konsekwencji. Obejmują one nieodpowiednie analizy potrzeb, brak studiów wykonalności i nieskoordynowane planowanie prowadzące do dodatkowych kosztów, których można było uniknąć.

W dojrzałym ekosystemie IT firmy wymiana przestarzałego systemu jest porównywalna z wymianą kolumny nośnej w już zbudowanym budynku. Nie wystarczy po prostu usunąć stary i zainstalować nowy - ważne jest, aby zrobić to w taki sposób, aby budynek pozostał stabilny, stropy się nie zawaliły, a cała komunikacja nadal działała. Na tym właśnie polega trudność: każdy błąd może mieć poważne konsekwencje dla całego systemu firmy.

Niemniej jednak twórcy dużych produktów ERP dla branży budowlanej nadal wykorzystują ilość napisanego kodu jako argument przemawiający za ich platformą. Na specjalistycznych konferencjach wciąż można usłyszeć sformułowania typu: "Odtworzenie takiego systemu zajęłoby 150 robocząt", mimo że większość funkcjonalności takich systemów kryje w sobie bazy danych i dość proste funkcje do pracy z tabelami, opakowane w specjalny, stały interfejs użytkownika. W praktyce objętość kodu "150 osobotat" staje się raczej obciążeniem niż przewagą konkurencyjną. Im więcej kodu - tym wyższy koszt wsparcia, tym trudniej dostosować się do nowych warunków i tym wyższy próg wejścia dla nowych programistów i klientów.

Wiele modułowych systemów budowlanych przypomina dziś niewygodne i przestarzałe "konstrukcje Frankenstein", w których każda nieostrożna zmiana może prowadzić do awarii. Każdy nowy moduł zwiększa złożoność i tak już przeciążonego systemu, zamieniając go w labirynt, który może zrozumieć tylko kilku specjalistów, co czyni go jeszcze bardziej trudnym w utrzymaniu i modernizacji.

Złożoność jest również realizowana przez samych programistów, którzy okresowo wstrzymują się z refaktoryzacją - rewizją architektury w celu uwzględnienia pojawienia się nowych technologii. Jednak nawet jeśli refaktoryzacja jest przeprowadzana regularnie, złożoność nieuchronnie rośnie. Architekci takich systemów przyzwyczajają się do rosnącej złożoności, ale dla nowych użytkowników i specjalistów staje się ona barierą nie do pokonania. W rezultacie cała wiedza jest skoncentrowana w rękach kilku programistów, a system przestaje być skalowalny. Na krótką metę tacy eksperci są przydatni, ale w dłuższej perspektywie stają się częścią problemu.

Organizacje będą nadal integrować "małe" dane z ich odpowiednikami Big Data i nierozsądne jest, aby ktokolwiek wierzył, że jedna aplikacja - bez względu na to, jak droga lub solidna - może obsłużyć wszystko [109].

- Phil Simon, gospodarz podcastu Conversations About Collaboration

Pojawia się uzasadnione pytanie: czy naprawdę potrzebujemy tak uciążliwych i zamkniętych systemów do obliczania kosztów i czasu pracy w formie tabel, skoro inne branże od dawna radzą sobie z podobnymi zadaniami za pomocą narzędzi analitycznych z otwartymi danymi i przejrzystą logiką?

Obecnie zamknięte platformy modułowe są nadal poszukiwane w branży budowlanej, głównie ze względu na specyfikę rozliczania kosztów (rys. 5.1-7). Takie systemy są często wykorzystywane do prowadzenia szarych lub nieprzejrzystych programów, umożliwiając ukrywanie rzeczywistych kosztów przed klientem. Jednak w miarę jak branża dojrzewa cyfrowo, przede wszystkim klienci, i wkracza w tak zwaną "erę uberyzacji", pośrednicy, a mianowicie firmy budowlane ze swoimi systemami ERP, stracą na znaczeniu w kalkulacjach czasu i kosztów. To na zawsze zmieni oblicze branży budowlanej. Przeczytaj więcej w ostatniej części książki oraz w rozdziale "Budownictwo 5.0: Jak zarabiać pieniądze, gdy nie można się już ukryć".

Tysiące starszych rozwiązań zgromadzonych w ciągu ostatnich 30 lat i tysiące osobolat zainwestowanych w ich rozwój zaczną szybko zniknąć. Przejście do otwartego, przejrzystego i elastycznego zarządzania danymi jest nieuniknione. Pytanie tylko, które firmy będą w stanie dostosować się do tych zmian, a które pozostaną zakładnikami starego modelu.

Podobna sytuacja ma miejsce w obszarze narzędzi CAD (BIM -), których dane wypełniają dziś parametry objętościowe jednostek projektowych w systemach ERP/PMIS -. Początkowo idea BIM (opracowana jeszcze w 2002 roku [110]) opierała się na koncepcji jednej zintegrowanej bazy danych, jednak w praktyce dzisiejsza praca z BIM wymaga całego zestawu specjalistycznego oprogramowania i formatów. To, co miało uprościć projektowanie i zarządzanie budową, przekształciło się w kolejną warstwę zastrzeżonych rozwiązań, które komplikują integrację i zmniejszają elastyczność biznesową.

Kolejne kroki: efektywne wykorzystanie danych projektowych

W tej części pokazaliśmy, w jaki sposób ustrukturyzowane dane stają się podstawą dokładnych obliczeń kosztów i harmonogramów projektów budowlanych. Automatyzacja procesów QTO, harmonogramowania i szacowania zmniejsza koszty pracy i znacznie poprawia dokładność wyników.

Podsumowując tę część, warto podkreślić główne praktyczne kroki, które pomogą zastosować omówione podejścia w codziennych zadaniach. Podejścia te są uniwersalne - przydadzą się zarówno w cyfrowej transformacji firmy, jak i w codziennej pracy specjalistów zajmujących się obliczeniami:

- Automatyzacja rutynowych obliczeń
 - Postaraj się znaleźć standardowe kosztorysy, które możesz wykorzystać w swojej pracy.
 - Przeanalizuj, jakie metody są wykorzystywane do kosztorysowania lub obliczania kosztów

prac lub procesów na placu budowy w Twoim kraju (Rys. 5.1-7).

- Jeśli pracujesz z systemem CAD - poznaj funkcje automatycznego wyodrębniania specyfikacji i danych QTO w oprogramowaniu CAD (BIM-).
- Użyj LLM do napisania szkicu kodu w celu zautomatyzowania obliczeń
- Tworzenie własnych narzędzi dla QTO**
 - Tworzenie skryptów lub tabel w celu automatyzacji liczenia woluminów
 - Standaryzacja kategorii i grup elementów w celu zapewnienia spójnego podejścia do oceny
 - Dokumentowanie metodologii obliczeń w celu zapewnienia powtarzalności wyników w nowych projektach.
- Włączenie różnych aspektów projektu do swojej pracy**
 - Jeśli pracujesz z systemami modułowymi, spróbuj wizualizować swoje procesy nie tylko jako diagramy lub wykresy, ale także na poziomie danych - zwłaszcza w formie tabel
 - Opanuj automatyczne łączenie danych pobranych z baz danych CAD z obliczeniami - za pomocą kodu Python wykorzystującego grupowanie, filtrowanie i agregację.
 - Tworzenie przejrzystych wizualizacji grup QTO w celu prezentowania złożonych informacji współpracownikom i klientom.

Kroki te pomogą zbudować zrównoważony system obliczeniowy oparty na automatyzacji i standaryzacji danych. Takie podejście poprawi dokładność i zmniejszy rutynę codziennych obliczeń.

Kolejne rozdziały koncentrują się na technicznych aspektach produktów CAD - (BIM-) oraz powodach, dla których bazy danych CAD są nadal trudne do zintegrowania z procesami biznesowymi firm. Jeśli nie jesteś zainteresowany historią wdrażania BIM w budownictwie, ewolucją narzędzi CAD i technicznymi aspektami pracy z tymi technologiami, możesz przejść bezpośrednio do siódmej części książki "Podejmowanie decyzji w oparciu o dane".



MAKSYMALNA WYGODA DZIĘKI WERSJI DRUKOWANEJ

Posiadasz darmową wersję cyfrową **Data-Driven Construction**. W celu wygodniejszej pracy i szybkiego dostępu do materiałów zalecamy zwrócenie uwagi na **wydanie drukowane**:



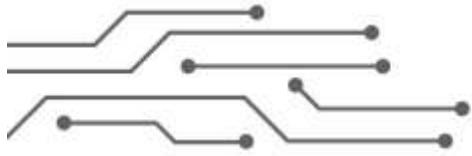
■ **Zawsze pod ręką:** książka w formacie drukowanym stanie się niezawodnym narzędziem pracy, pozwalającym szybko znaleźć i wykorzystać odpowiednie wizualizacje i diagramy w każdej sytuacji roboczej.

■ **Wysokiej jakości ilustracje:** wszystkie obrazy i grafiki w drukowanym wydaniu są prezentowane w maksymalnej jakości.

■ **Szybki dostęp do informacji:** łatwa nawigacja, możliwość tworzenia notatek, zakładek i pracy z książką w dowolnym miejscu.

Kupując pełną drukowaną wersję książki, otrzymujesz wygodne narzędzie do wygodnej i efektywnej pracy z informacją: możliwość szybkiego korzystania z materiałów wizualnych w codziennych zadaniach, szybkiego wyszukiwania potrzebnych schematów i robienia notatek. Ponadto zakup wspiera rozpowszechnianie otwartej wiedzy.

Zamów drukowaną wersję książki na stronie: datadrivenconstruction.io/books



V CZĘŚĆ

CAD I BIM: MARKETING, RZECZYWISTOŚĆ I PRZYSZŁOŚĆ DANYCH PROJEKTOWYCH W BUDOWNICTWIE

Szósta część książki przedstawia krytyczną analizę ewolucji technologii CAD i BIM oraz ich wpływu na procesy zarządzania danymi w budownictwie. Śledzi historyczną transformację koncepcji BIM od pierwotnej idei zintegrowanej bazy danych do obecnych konstrukcji marketingowych promowanych przez producentów oprogramowania. Oceniono wpływ zastrzeżonych formatów i zamkniętych systemów na efektywność obsługi danych projektowych i ogólną wydajność branży budowlanej. Szczegółowo przeanalizowano problemy kompatybilności różnych systemów CAD i trudności ich integracji z procesami biznesowymi firm budowlanych. Omówiono aktualne trendy w kierunku uproszczonych otwartych formatów danych, takich jak USD, oraz ich potencjalny wpływ na branżę. Przedstawiono alternatywne podejścia do wydobywania informacji z zamkniętych systemów, w tym techniki inżynierii odwrotnej. Przeanalizowano perspektywy zastosowania sztucznej inteligencji i uczenia maszynowego do automatyzacji procesów projektowania i analizy danych w budownictwie. Sformułowano prognozy rozwoju technologii projektowych zorientowanych na rzeczywiste potrzeby użytkowników, a nie interesy producentów oprogramowania.

ROZDZIAŁ 5.1.

POJAWIENIE SIĘ KONCEPCJI BIM- W BRANŻY BUDOWLANEJ

Pierwotnie ta szósta część, poświęcona CAD (BIM), nie została uwzględniona w pierwszej wersji książki. Tematy zastrzeżonych formatów, geometrycznych jąder i zamkniętych systemów są zbyt techniczne, przeładowane szczegółami i pozornie bezużyteczne dla tych, którzy chcą po prostu zrozumieć, jak pracować z danymi. Jednak informacje zwrotne i prośby o dodanie wyjaśnień do pierwszej wersji książki pokazały, że bez zrozumienia złożoności wewnętrznego działania systemów CAD, jąder geometrii, różnorodności formatów i niekompatybilnych schematów przechowywania tych samych danych, niemożliwe jest prawdziwe zrozumienie, dlaczego koncepcje promowane przez dostawców często utrudniają pracę z informacjami i utrudniają przejście do otwartego projektowania parametrycznego. Dlatego właśnie ta część zajęła swoje własne miejsce w strukturze książki. Jeśli CAD (BIM) nie jest dla Ciebie priorytetem, możesz przejść od razu do następnej części - "CZĘŚĆ VII: Podejmowanie decyzji w oparciu o dane, analityka, automatyzacja i uczenie maszynowe".

Historia pojawienia się BIM i otwartego BIM jako koncepcji marketingowych dostawców CAD-

Wraz z pojawieniem się danych cyfrowych w latach 90-tych, technologia komputerowa została wprowadzona nie tylko w procesach biznesowych, ale także w procesach projektowych, co doprowadziło do powstania takich koncepcji jak CAD (systemy projektowania wspomaganego komputerowo), a później BIM (modelowanie informacji budynku).

Jednak, jak każda innowacja, nie są one punktem końcowym rozwoju. Koncepcje takie jak BIM stały się ważnym kamieniem milowym w historii branży budowlanej, ale przedżej czy później mogą ustąpić miejsca lepszym narzędziom i podejściom, które lepiej sprostają wyzwaniom przyszłości.

Koncepcja BIM, która pojawiła się w 2002 roku, przytłoczona wpływem dostawców CAD i zdezorientowana złożonością własnej implementacji, może nie dożyć trzydziestej rocznicy, niczym gwiazda rocka, która rozblysła jasno, ale szybko zgasła. Powód jest prosty: wymagania naukowców zajmujących się danymi zmieniają się szybciej niż dostawcy CAD mogą się do nich dostosować.

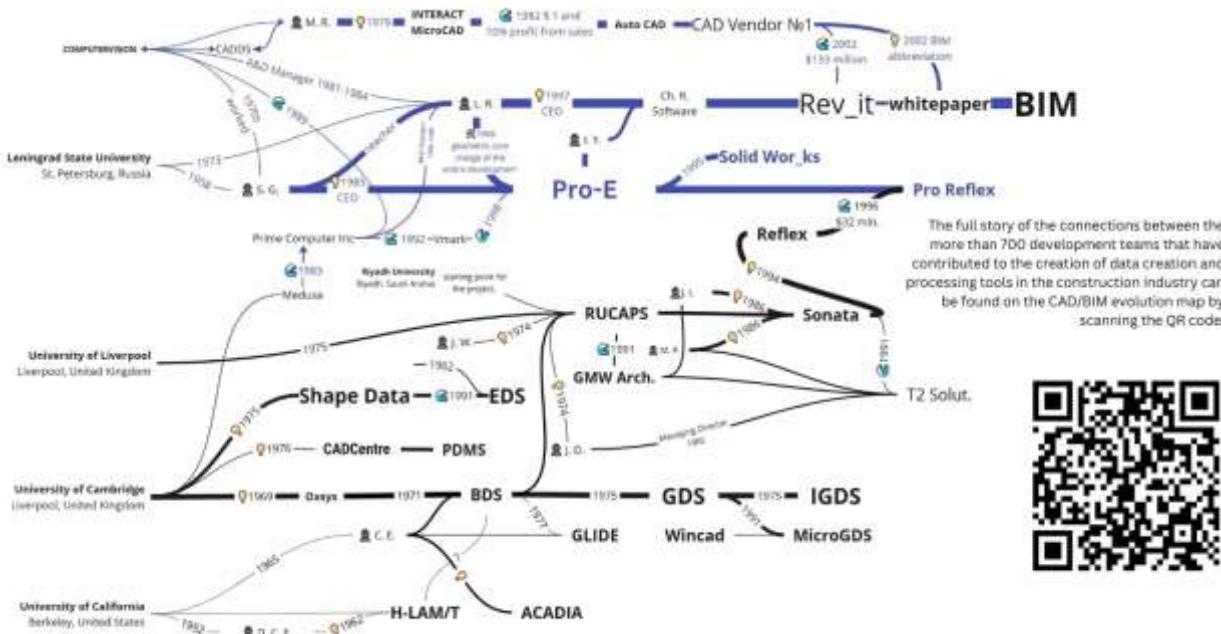
W obliczu braku wysokiej jakości danych, dzisiejsi profesjonalisci z branży budowlanej domagają się interoperacyjności międzyplatformowej i dostępu do otwartych danych z projektów CAD- w celu uproszczenia ich analizy i przetwarzania. Złożoność danych CAD i myjące przetwarzanie danych CAD ma negatywny wpływ na wszystkich zaangażowanych w proces budowlany: projektantów, kierowników projektów, pracowników budowlanych na placu budowy i ostatecznie na klienta.

Zamiast pełnowartościowego zbioru danych do pracy, klient i inwestor otrzymują dziś kontenery w formatach CAD-, które wymagają złożonego jądra geometrycznego, zrozumienia schematów danych, corocznie aktualizowanej dokumentacji API i specjalistycznego oprogramowania CAD (BIM) do pracy z danymi. Jednocześnie znaczna część danych projektowych pozostaje niewykorzystana.

W dzisiejszym świecie projektowania i budownictwa złożoność dostępu do danych CAD prowadzi do nadmiernej inżynierii zarządzania projektami. Średnie i duże firmy pracujące z danymi CAD lub opracowujące rozwiązania BIM są albo zmuszone do utrzymywania bliskich relacji z dostawcami rozwiązań CAD w celu uzyskania dostępu do danych za pośrednictwem interfejsów API, albo do omijania ograniczeń dostawców CAD poprzez stosowanie drogich konwerterów SDK do inżynierii odwrotnej, w celu uzyskania otwartych danych [75].

Podejście oparte na zastrzeżonych danych jest przestarzałe i nie spełnia już wymagań dzisiejszego środowiska cyfrowego. Przyszłość podzieli firmy na dwa rodzaje: te, które efektywnie wykorzystują otwarte dane oraz te, które opuszczą rynek.

Koncepcja BIM (Building Information Modeling), pojawiła się w branży budowlanej wraz z publikacją jednego z głównych dostawców CAD - Whitepaper BIM [54] w 2002 roku i, uzupełniona koncepcją inżynierii mechanicznej BOM (Bills of Materials), wywodzi się z parametrycznego podejścia do tworzenia i przetwarzania danych projektowych (rys. 6.1-1). Parametryczne podejście do tworzenia i przetwarzania danych projektowych było jednym z pierwszych wdrożonych w systemie Pro-E do projektowania inżynierii mechanicznej (MCAD). System ten stał się prototypem [111] dla wielu nowoczesnych rozwiązań CAD, w tym tych stosowanych obecnie w branży budowlanej.



Modelowanie informacji o budynku to strategia..... [nazwa dostawcy CAD] w celu zastosowania technologii informatycznych w przemyśle budowlanym.

- BIM Whitepaper, 2002 [60]

W połowie XXI wieku "badacze" zaczęli łączyć koncepcję BIM- opublikowaną przez dostawcę CAD- w 2002 roku z wcześniejszymi pracami naukowymi, takimi jak BDS Charlesa Eastmana, które stały się podstawą systemów takich jak GLIDE, GBM, BPM, RUCAPS. W swojej pionierskiej pracy Building Description System (1974) Charles Eastman położył teoretyczne podwaliny nowoczesnego modelowania informacji. Termin "baza danych" pojawia się w jego pracy 43 razy (rys. 6.1-2) - częściej niż jakikolwiek inny, z wyjątkiem słowa "budynek".

Cluczową ideą Eastmana było to, że wszystkie informacje o budynku - od geometrii po właściwości elementów i ich wzajemne powiązania - powinny być przechowywane w jednej ustrukturyzowanej bazie danych. To właśnie z tej bazy danych można automatycznie generować i analizować rysunki, specyfikacje, obliczenia i zgodność z przepisami. Eastman wyraźnie skrytykował rysunki jako przestarzałą i zbędną metodę komunikacji, wskazując na powielanie informacji, problemy z aktualizacją i potrzebę ręcznej aktualizacji po wprowadzeniu zmian. Zamiast tego zaproponował pojedynczy model cyfrowy w bazie danych, w którym każda zmiana jest wprowadzana raz i automatycznie odzwierciedlana we wszystkich widokach.

Warto zauważyć, że w swojej koncepcji Eastman nie umieścił wizualizacji w centrum. Centralne miejsce w jego systemie zajmowały informacje: parametry, relacje, atrybuty, możliwości analizy i automatyzacji. Rysunki w jego rozumieniu były tylko jedną z form wyświetlania danych z bazy danych, a nie głównym źródłem informacji projektowych.

W pierwszym Whitepaper na temat BIM od wiodącego dostawcy CAD, wyrażenie "baza danych" było używane tak często, jak w BDS Charlesa Eastmana - 23 razy [60] na siedmiu stronach i było jednym z najpopularniejszych słów w dokumencie po "Budynku", "Informacji", "Modelowaniu" i "Projektowaniu". Jednak do 2003 roku termin "baza danych" pojawił się tylko dwa razy w podobnych dokumentach [61], a pod koniec 2000 roku temat baz danych praktycznie zniknął z dyskusji na temat danych projektowych. W rezultacie koncepcja "pojedynczej zintegrowanej bazy danych do analizy wizualnej i ilościowej" nigdy nie została w pełni zrealizowana.

W ten sposób branża budowlana przeszła od progresywnej koncepcji BDS Charlesa Eastmana z naciskiem na bazy danych i pomysły Samuela Geisberga dotyczące automatycznej aktualizacji danych projektowych z baz danych w produkcie inżynierii mechanicznej Pro-E (poprzednik popularnego CAD -rozwiązań stosowane obecnie w budownictwie) do obecnie sprzedawanego BIM, w którym zarządzanie danymi za pośrednictwem baz danych jest ledwo wspomniane, mimo że była to koncepcja stojąca za oryginalną teorią.

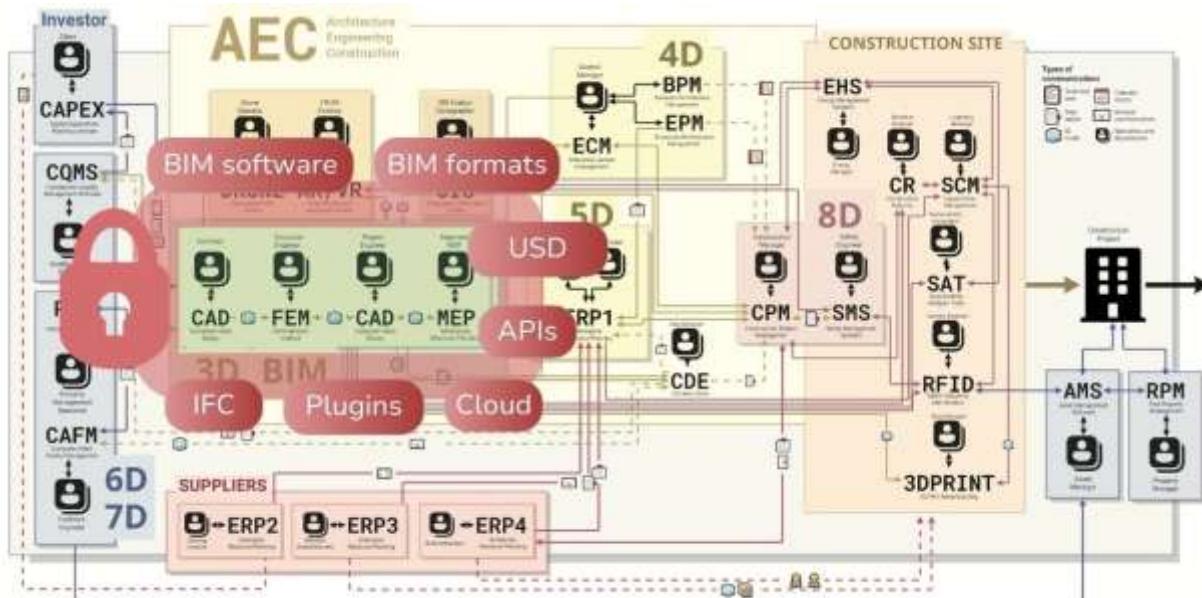
zastrzeżonych interfejsów API.

- **Dane należą do sprzedawców, a nie użytkowników.** Informacje o projektach są zablokowane w zastrzeżonych formatach lub usługach w chmurze, zamiast być dostępne w otwartych i niezależnych formatach.

Inżynierowie projektanci i kierownicy projektów często nie mają dostępu do bazy danych CAD-systems, ani do formatu, w którym przechowywane są ich własne dane projektowe. Uniemożliwia to szybką weryfikację informacji lub sformułowanie wymagań dotyczących struktury i jakości danych (rys. 6.1-3). Dostęp do takich danych wymaga całego zestawu wyspecjalizowanych programów połączonych za pośrednictwem interfejsów API i wtyczek, co prowadzi do nadmiernej biurokratyzacji procesów w branży budowlanej. Tymczasem dane te są jednocześnie wykorzystywane przez dziesiątki systemów informatycznych i setki specjalistów.

Musimy być w stanie zarządzać wszystkimi tymi danymi [CAD (BIM)], przechowywać je cyfrowo i sprzedawać oprogramowanie do zarządzania cyklem życia i procesami, ponieważ na każdego inżyniera [projektanta], który tworzy coś [w oprogramowaniu CAD], przypada dziesięć osób, które pracują z tymi danymi" [41].

- CEO CAD - dostawcy, który stworzył koncepcję BIM, 2005.



Rys. 6.1-3 CAD- (BIM-) bazy danych pozostają jednym z ostatnich zamkniętych systemów dla działów IT i menedżerów danych w ekosystemie biznesu budowlanego.

Kiedy staje się oczywiste, że BIM jest raczej środkiem komercjalizacji baz danych niż pełnoprawnym narzędziem do zarządzania bazami danych, pojawia się logiczne pytanie: w jaki sposób możemy odzyskać kontrolę nad danymi? Odpowiedzią jest wykorzystanie otwartych struktur danych, w których użytkownik, a nie dostawca oprogramowania, staje się właścicielem informacji.

Użytkownicy i twórcy rozwiązań w branży budowlanej, podobnie jak ich odpowiednicy w innych branżach, nieuchronnie odejdą od niejasnej terminologii dostawców oprogramowania, która dominowała w ciągu ostatnich 30 lat, koncentrując się na kluczowych aspektach cyfryzacji - "danych" i "procesach".

Jeszcze w późnych latach 80. kluczowy obszar rozwoju cyfrowego w budownictwie był przedstawiany jako kwestia dostępu do danych i zarządzania informacjami projektowymi. Z czasem jednak punkt ciężkości uległ zmianie. Zamiast opracowywać przejrzyste i dostępne podejście do pracy z danymi, aktywnie promowano format IFC i otwartą koncepcję BIM jako próby odwrócenia uwagi specjalistów od tematów zarządzania bazami danych projektów.

Pojawienie się otwartego formatu IFC w branży budowlanej

Tak zwany otwarty format IFC (Industry Foundation Classes) jest pozycjonowany jako standard zapewniający interoperacyjność między różnymi systemami CAD (BIM -). Jego rozwój został przeprowadzony w ramach organizacji stworzonych i kontrolowanych przez głównych dostawców CAD. W oparciu o format IFC dwie firmy CAD opracowały w 2012 r. koncepcję marketingową OPEN BIM[63].

IFC (Industry Foundation Classes) to otwarty standard wymiany danych w branży budowlanej, zaprojektowany w celu zapewnienia interoperacyjności między różnymi systemami CAD - (BIM-).

Koncepcja Open BIM - obejmuje pracę z informacjami z baz danych CAD i wymianę informacji między systemami za pośrednictwem otwartego formatu wymiany danych CAD - IFC.

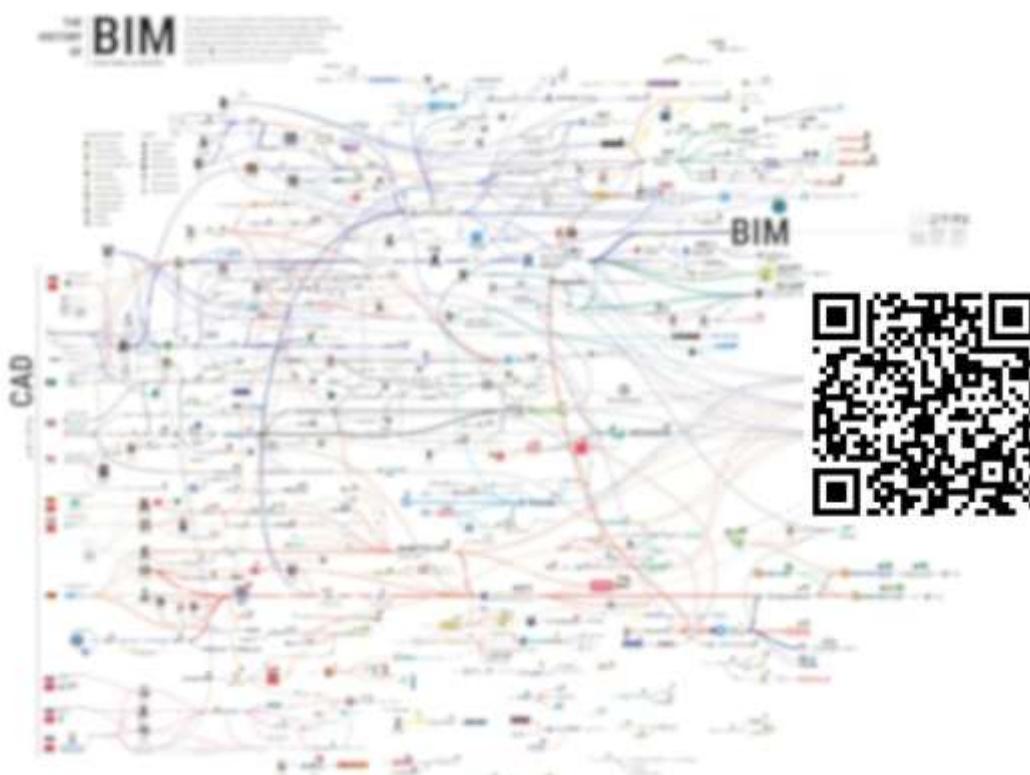
Program Open BIM to kampania marketingowa zainicjowana przez... [1 dostawcę CAD],... [2 dostawcę CAD] i inne firmy w celu zachęcenia i ułatwienia globalnej skoordynowanej promocji koncepcji OPEN BIM w całej branży AEC, ze spójną komunikacją i wspólnym brandingiem dostępnym dla uczestników programu.

- Ze strony internetowej dostawcy CAD, Program OPEN BIM, 2012 [113]

IFC został zaadaptowany przez Uniwersytet Techniczny w Monachium z formatu inżynierii mechanicznej STEP pod koniec lat 80-tych, a następnie zarejestrowany przez dużą firmę projektową i głównego dostawcę CAD w celu utworzenia IAI (Industry Alliance for Interoperability) w 1994 roku [114] (rys. 6.1-4). Format IFC został opracowany w celu zapewnienia interoperacyjności między różnymi systemami CAD i został oparty na zasadach określonych w formacie inżynierii mechanicznej STEP, który z kolei wyłonił się z formatu IGES stworzonego w 1979 roku przez grupę użytkowników i sprzedawców CAD przy wsparciu NIST (The National

Institute of Standards and Technology) i Departamentu Obrony USA [115].

Jednak złożona struktura IFC, jego ścisła zależność od jądra geometrycznego, a także rozbieżności w implementacji formatu przez różne rozwiązania programowe doprowadziły do wielu problemów w jego praktycznym zastosowaniu. Podobne trudności - utrata szczegółów, ograniczenie dokładności i konieczność stosowania formatów pośrednich - były wcześniej napotykane przez specjalistów inżynierii mechanicznej podczas pracy z formatami IGES i STEP, z których wyłonił się IFC.



Rys. 6.1-4 Mapa powiązań między zespołami programistów a produktami CAD (BIM) [116].

W 2000 r. ten sam dostawca CAD, który zarejestrował format IFC i stworzył organizację IAI (później bS), publikuje białą księgę "Zintegrowane projektowanie i produkcja: korzyści i uzasadnienie" [65]. W dokumencie podkreślono znaczenie zachowania pełnej ziarnistości danych podczas wymiany między programami w ramach tego samego systemu, bez użycia neutralnych formatów, takich jak IGES, STEP [identyczny z IFC]. Zamiast tego zaproponowano, aby aplikacje miały bezpośredni dostęp do podstawowej bazy danych CAD, aby zapobiec utracie dokładności informacji.

W 2002 roku ten sam dostawca CAD kupuje parametryczny produkt BOM (rys. 3.1-18, więcej szczegółów w części trzeciej) i na jego podstawie tworzy koncepcję BIM. W rezultacie tylko zamknięte formaty CAD lub format IFC (STEP) są używane do wymiany danych projektu budowlanego, o których ograniczeniach pisał sam sprzedawca CAD w 2000 roku, który wprowadził ten format do branży budowlanej.

Szczegółowa historia interakcji ponad 700 zespołów programistycznych zajmujących się tworzeniem i

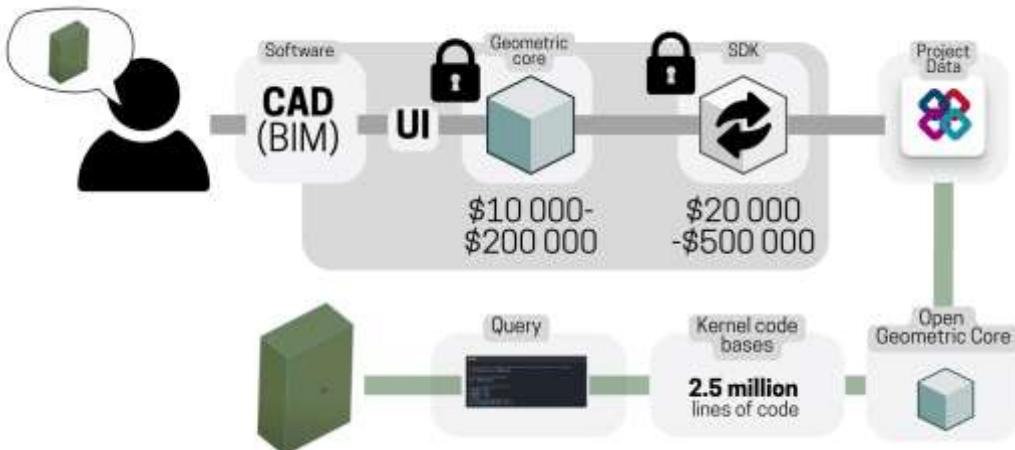
przetwarzaniem danych budowlanych została przedstawiona na mapie "Ewolucja CAD (BIM)" [116]. [116].

Otwarta forma IFC składa się z opisu geometrycznego elementów projektu i opisu metainformacji. Do reprezentowania geometrii w formacie IFC stosowane są różne metody, takie jak CSG i Swept Solids; jednak reprezentacja parametryczna BREP stała się wiodącym standardem przesyłania geometrii elementów w formacie IFC, ponieważ format ten jest obsługiwany podczas eksportowania z programów CAD- (BIM-) i umożliwia potencjalną edycję elementów podczas importowania IFC z powrotem do programów CAD.

Problem z formatem IFC w zależności od rdzenia geometrycznego

W większości przypadków, gdy geometria w IFC jest zdefiniowana parametrycznie (BREP), niemożliwe staje się wizualizowanie lub pobieranie właściwości geometrycznych, takich jak objętość lub powierzchnia elementów projektu, przy użyciu tylko pliku IFC, ponieważ do pracy i wizualizacji geometrii w tym przypadku wymagane jest jądro geometrii (Rys. 6.1-5), którego początkowo brakuje.

Jądro geometrii to komponent oprogramowania, który zapewnia podstawowe algorytmy tworzenia, edycji i analizy obiektów geometrycznych w CAD (CAD), BIM i innych aplikacjach inżynierskich. Odpowiada za konstrukcję geometrii 2D i 3D, a także za operacje na niej, takie jak: operacje boolowskie, wygładzanie, przecięcia, transformacje i wizualizacje.

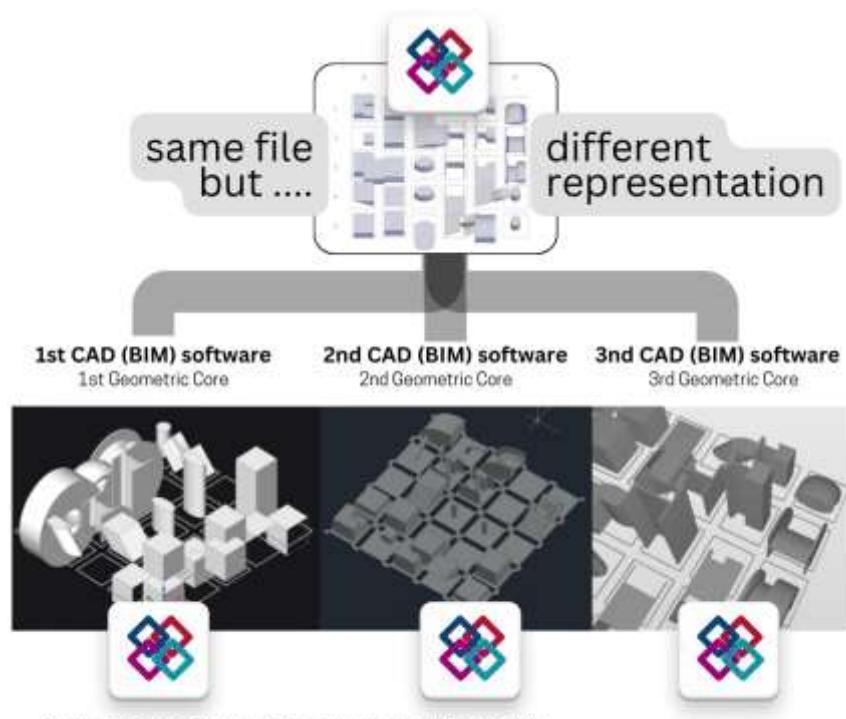


Rys. 6.1-5 Tworzenie geometrii za pomocą oprogramowania CAD- odbywa się obecnie za pomocą zastrzeżonych jąder geometrii i zestawów SDK, które często nie są własnością dostawców CAD.

Każdy program CAD i wszystkie programy pracujące z formatami parametrycznymi lub IFC mają własne lub zakupione jądro geometryczne. I jeśli z prymitywnymi elementami w formacie IFC -BREP nie może być problemów, a w programach z różnymi jądrami geometrycznymi elementy te mogą być wyświetlane podobnie, ale oprócz problemów z różnymi silnikami jąder geometrycznych, istnieje wystarczająco dużo elementów, które mają swoje własne cechy szczególne dla prawidłowego wyświetlania. Problem ten został szczegółowo omówiony w międzynarodowym badaniu "A reference study of IFC software support" opublikowanym w 2019 roku [117].

Te same znormalizowane zbiory danych dają sprzeczne wyniki, przy czym znaleziono niewiele wspólnych wzorców, a także stwierdzono poważne problemy we wspieraniu standardu [IFC], prawdopodobnie ze względu na bardzo dużą złożoność standardowego modelu danych. Częściowo winę za to ponoszą same standardy, ponieważ często pozostawiają one pewne szczegóły niezdefiniowane, z dużymi stopniami swobody i różnymi możliwymi interpretacjami. Pozwalają one na dużą złożoność w organizacji i przechowywaniu obiektów, co nie sprzyja skutecznemu uniwersalnemu zrozumieniu, unikalnym implementacjom i spójnemu modelowaniu danych [117].

- Badanie referencyjne wsparcia oprogramowania IFC, 2021 r.



Rys. 6.1-6 Różne jądra geometryczne dają różne reprezentacje tej samej geometrii opisanej parametrycznie (na podstawie [117]).

Prawidłowe zrozumienie "niektórych przepisów" jest dostępne dla płatnych członków specjalnych organizacji, które rozwijają IFC. W konsekwencji każdy, kto chce uzyskać dostęp do ważnej wiedzy na temat pewnych cech IFC, będzie próbował współpracować z dużymi dostawcami CAD- lub osiągnąć jakościowe rozważenie cech poprzez własne badania

Natknąłeś się na pytanie dotyczące importowania i eksportowania danych za pośrednictwem formatu IFC i zapytałeś innych sprzedawców: "Dlaczego w pliku IFC jest informacja o parametrycznym przenoszeniu pomieszczeń? Otwarta specyfikacja nic o tym nie mówi". Odpowiedź od "bardziej kompetentnych" europejskich sprzedawców: "Tak, nie jest to powiedziane, ale jest to dozwolone".

- Z wywiadu z deweloperem CAD 2021 [118]

IFC opisuje geometrię poprzez parametryczne prymitywy, ale nie zawiera wbudowanego jądra - jego rolę pełni program CAD, który kompliuje geometrię poprzez jądro geometrii. Jądro geometrii wykonuje obliczenia matematyczne i definiuje przecięcia, podczas gdy IFC dostarcza jedynie dane do jego interpretacji. Jeśli IFC zawiera nieprawidłowe ściany, różne programy z różnymi jądrami geometrii mogą je zignorować lub wygenerować błędy, w zależności od jądra.

W rezultacie, aby pracować z formatem IFC należy odpowiedzieć sobie na główne pytanie, na które trudno znaleźć jednoznaczną odpowiedź - jakiego narzędzia, z jakim jądem geometrycznym należy użyć, aby uzyskać jakość danych, jaką pierwotnie posiadał projekt w programie CAD, z którego uzyskano IFC?

Kwestie jakości danych i złożoność formatu IFC nie pozwalają na bezpośrednie wykorzystanie danych projektowych do automatyzacji procesów, analizy i przetwarzania danych, co często prowadzi deweloperów do nieuniknionej konieczności korzystania z zamkniętych rozwiązań CAD z "jakościowym" dostępem do danych[63], o czym pisał sam sprzedawca, który zarejestrował IFC w 1994 roku [65].

Wszystkie osobliwości mapowania i generowania parametrów IFC w jądrze geometrii mogą być realizowane tylko przez duże zespoły programistów, którzy mają doświadczenie w pracy z jądrami geometrii. Dlatego obecna praktyka osobliwości i złożoności formatu IFC jest korzystna przede wszystkim dla dostawców CAD- i ma wiele wspólnego ze strategią dużych producentów oprogramowania "adoptuj, rozszerzaj, niszcz", gdy rosnąca złożoność standardu faktycznie tworzy bariery dla małych graczy rynkowych [94].

Strategia dużych dostawców w takiej strategii może polegać na dostosowywaniu otwartych standardów, dodawaniu zastrzeżonych rozszerzeń i funkcji, aby stworzyć zależność użytkownika od ich produktów, a następnie wyprzeć konkurencję.

Format IFC, mający być uniwersalnym pomostem pomiędzy różnymi systemami CAD- (BIM-), w rzeczywistości pełni rolę wskaźnika problemów z kompatybilnością pomiędzy geometrycznymi rdzeniami różnych platform CAD, podobnie jak format STEP, z którego pierwotnie się wyłonił.

W rezultacie, obecnie pełna i wysokiej jakości implementacja ontologii IFC jest wykonalna dla dużych dostawców CAD, którzy mogą zainwestować znaczne zasoby w celu wsparcia wszystkich podmiotów i ich mapowania do własnego wewnętrznego rdzenia geometrii, który nie istnieje dla IFC jako standardu. Duzi dostawcy mają również możliwość koordynowania między sobą szczegółów technicznych funkcji, które

mogą nie być dostępne nawet dla najbardziej aktywnego uczestnika organizacji zajmujących się rozwojem formatu IFC.

Dla małych niezależnych zespołów i projektów open-source, dążących do wspierania rozwoju interoperacyjnych formatów, brak własnego jądra geometrii staje się poważnym problemem. Bez niego praktycznie niemożliwe jest uwzględnienie wszystkich subtelności i niuansów związanych z międzyplatformową wymianą danych.

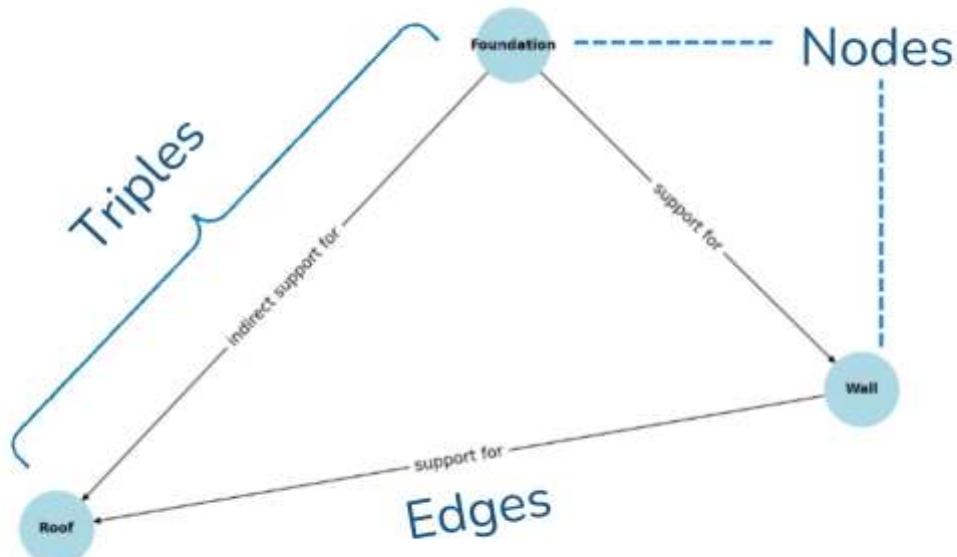
Wraz z rozwojem formatu parametrycznego IFC i otwartej koncepcji BIM, w branży budowlanej nasiliły się dyskusje na temat roli ontologii i semantyki w zarządzaniu danymi i procesami.

Pojawienie się w konstrukcji tematu semantyki i ontologii

Dzięki ideom semantycznego Internetu pod koniec 90. i wysiłkom organizacji zaangażowanych w rozwój formatu IFC, semantyka i ontologie stały się jednymi z kluczowych elementów standaryzacji omawianej w branży budowlanej do połowy 2020 roku.

Technologie semantyczne to ujednolicenie, standaryzacja i modyfikacja dużych tablic heterogenicznych danych oraz wdrażanie złożonych wyszukiwań.

OWL (Web Ontology Language), reprezentowany jako grafy RDF -triplets (Resource Description Framework) (rys. 6.1-7), służy do przechowywania danych semantycznych. OWL odnosi się do modeli danych grafowych, których typy omówiliśmy bardziej szczegółowo w rozdziale "Modele danych: relacje danych i relacje między elementami".



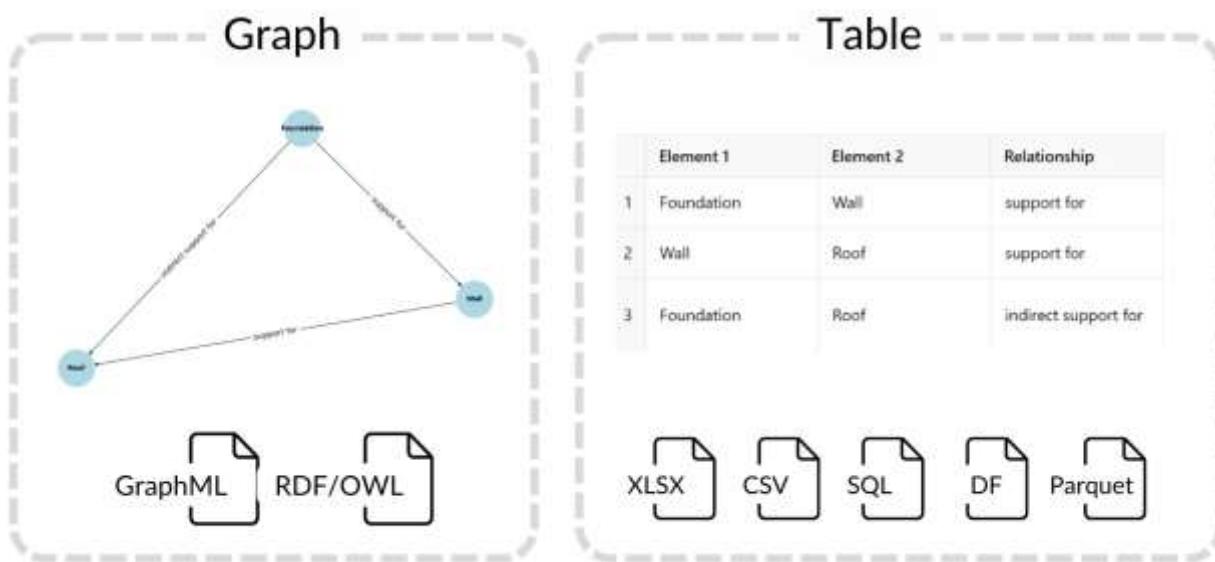
Rys. 6.1-7 Model danych RDF: węzły, krawędzie i trójkę ilustrujące relacje między elementami składowymi.

Teoretycznie, logiczne wnioskowanie risonerów (oprogramowanie do automatycznego wnioskowania logicznego) pozwala na wyprowadzanie nowych stwierdzeń z ontologii. Na przykład, jeśli ontologia budynku rejestruje, że "fundament jest podporą dla ściany" i "ściana jest podporą dla dachu" (rys. 6.1-7), risoner jest w stanie automatycznie wywnioskować, że "fundament jest podporą dla dachu".

Taki mechanizm jest przydatny do optymalizacji analizy danych, ponieważ pozwala uniknąć wyraźnego określania wszystkich zależności. Nie tworzy on jednak nowej wiedzy, a jedynie identyfikuje i strukturyzuje już znane fakty.

Semantyka nie tworzy nowego znaczenia ani wiedzy per se i nie jest w tym aspekcie lepsza od innych technologii przechowywania i przetwarzania danych. Reprezentowanie danych z relacyjnych baz danych jako trójkę nie czyni ich bardziej znaczącymi. Zastąpienie tabel strukturami grafowymi może być przydatne do ujednolicenia modeli danych, łatwego wyszukiwania i bezpiecznej edycji, ale nie sprawia, że dane stają się "mądrzejsze" - komputer nie zaczyna lepiej rozumieć ich zawartości.

Logiczne relacje w danych mogą być organizowane bez złożonych technologii semantycznych (rys. 6.1-8). Tradycyjne relacyjne bazy danych (SQL), a także formaty CSV lub XLSX umożliwiają tworzenie podobnych zależności. Na przykład w kolumnowej bazie danych można dodać pole "podpora dachu" i automatycznie powiązać dach z fundamentem podczas tworzenia ściany. Podejście to jest realizowane bez użycia RDF, OWL, grafów czy risonerów, pozostając prostym i wydajnym rozwiązaniem do przechowywania i analizy danych.



Rys. 6.1-8 Porównanie modeli danych grafu i tabeli dla reprezentowania tych samych relacji logicznych.

Decyzja wielu dużych firm budowlanych i organizacji zajmującej się rozwojem formatu IFC [94], aby podążać za koncepcją sieci semantycznej, która wydawała się obiecująca pod koniec lat 90-tych, miała znaczący wpływ na rozwój standardów w branży budowlanej.

Paradoks polega jednak na tym, że sama koncepcja sieci semantycznej, pierwotnie przeznaczona dla Internetu, nie została szeroko przyjęta nawet w swoim rodzimym środowisku. Pomimo rozwoju RDF i OWL, w pełni rozwinięta sieć semantyczna nie pojawiła się w swojej pierwotnej koncepcji, a jej stworzenie jest już mało prawdopodobne.

Dlaczego technologie semantyczne nie spełniają oczekiwania branży budowlanej?

Inne branże borykają się z ograniczeniami technologii wykorzystania semantyki. W branży gier próby opisania obiektów gier i ich interakcji za pomocą ontologii okazały się nieskuteczne ze względu na dużą dynamikę zmian. W rezultacie preferowano prostsze formaty danych, takie jak XML i JSON, wraz z rozwiązaniami algorytmicznymi. Podobna sytuacja miała miejsce w sektorze nieruchomości: ze względu na regionalne różnice w terminologii i częste zmiany na rynku, stosowanie ontologii okazało się zbyt skomplikowane, podczas gdy proste bazy danych i standardy, takie jak RETS [119], lepiej radziły sobie z wyzwaniami związanymi z wymianą danych.

Trudności techniczne, takie jak złożoność znaczników, wysoka pracochłonność i niska motywacja programistów, spowolniły przyjęcie sieci semantycznej i w innych sektorach gospodarki. RDF (Resource Description Framework) nie stał się masowym standardem, a ontologie okazały się zbyt złożone i ekonomicznie nieuzasadnione.

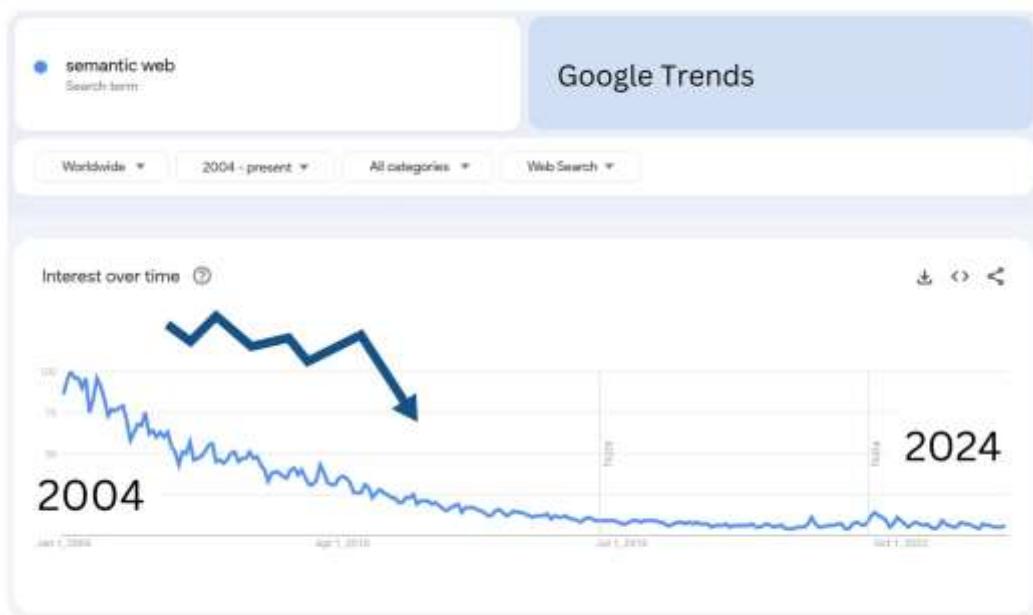
W rezultacie ambitny pomysł stworzenia globalnej sieci semantycznej nie powiodł się. Chociaż niektóre elementy technologii, takie jak ontologie i SPARQL, znalazły zastosowanie w rozwiązańach korporacyjnych, pierwotny cel stworzenia jednej kompleksowej struktury danych nie został osiągnięty.

Koncepcja Internetu, w którym komputery są w stanie nadać sens treści, okazała się technicznie trudna i komercyjnie nieopłacalna. Dlatego też firmy, które wspierały ten pomysł, ostatecznie ograniczyły jego wykorzystanie do pojedynczych przydatnych narzędzi, pozostawiając RDF i OWL dla wysoce wyspecjalizowanych potrzeb korporacyjnych, a nie dla Internetu jako całości. Analiza trendów Google (Rys. 6.1-9) w ciągu ostatnich 20 lat sugeruje, że nie ma już perspektyw dla sieci semantycznej.

Nie ma potrzeby niepotrzebnego mnożenia bytów. Jeśli istnieje kilka logicznie spójnych wyjaśnienia danego zjawiska, które wyjaśniają je równie dobrze, należy, przy wszystkich innych rzeczach równych, preferować najprostsze z nich.

- Brzytwa

Pojawia się tutaj logiczne pytanie: po co w ogóle używać trójkę, pionów i SPARQL w konstrukcji, skoro można przetwarzać dane za pomocą popularnych zapytań strukturalnych (SQL, Pandas, Apache®)? W zastosowaniach korporacyjnych SQL jest standardem pracy z bazami danych. SPARQL, przeciwnie, wymaga złożonych struktur grafowych i specjalistycznego oprogramowania i, zgodnie z trendami w Google, nie przyciąga zainteresowania programistów.



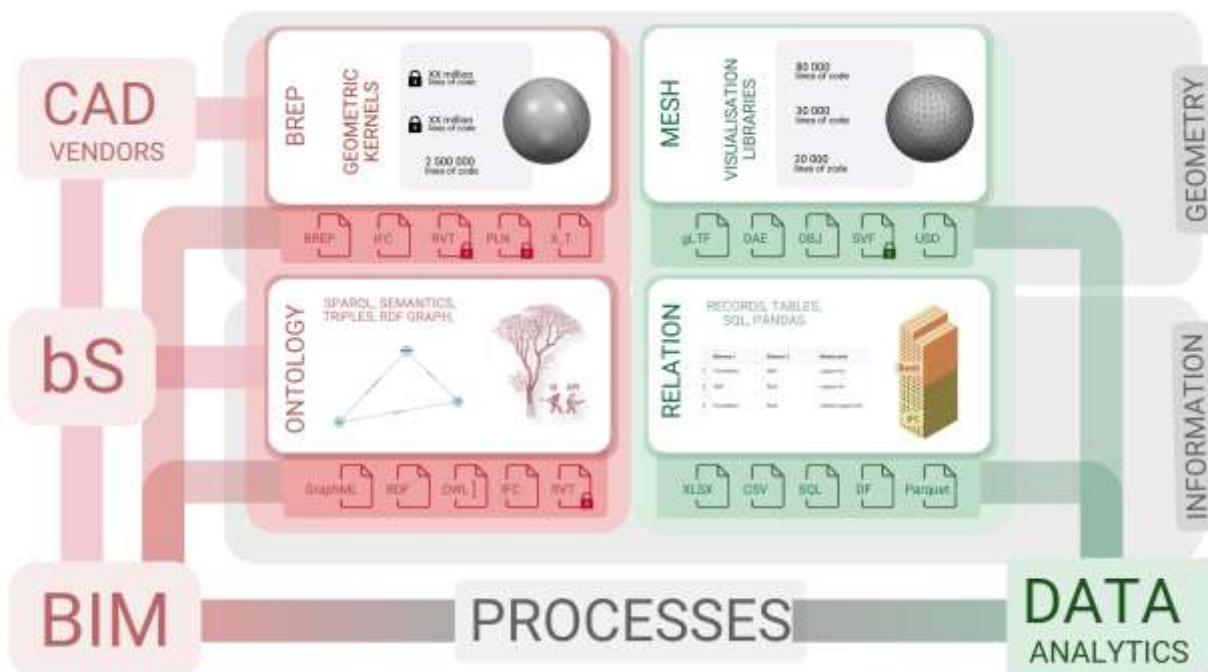
Rys. 6.1-9 Zainteresowanie zapytaniami "internet semantyczny" według statystyk Google.

Grafowe bazy danych i drzewa klasyfikacyjne mogą być przydatne w niektórych przypadkach, ale ich zastosowanie nie zawsze jest uzasadnione w przypadku większości codziennych zadań. W rezultacie tworzenie grafów wiedzy i korzystanie z technologii sieci semantycznej ma sens tylko wtedy, gdy konieczne jest ujednolicenie danych z różnych źródeł lub realizacja złożonych wniosków logicznych.

Przejście od tabel do grafowych modeli danych usprawnia wyszukiwanie i ujednolica przepływ informacji, ale nie sprawia, że dane stają się bardziej znaczące dla maszyn. Pytanie nie brzmi, czy technologie semantyczne powinny być stosowane, ale gdzie naprawdę robią różnicę.

Przed wdrożeniem ontologii, semantyki i grafowych baz danych w swojej firmie, dowiedz się, które firmy już z powodzeniem stosują te technologie, a gdzie poniosły porażkę.

Pomimo ambitnych oczekiwaniń, technologie semantyczne nigdy nie stały się uniwersalnym rozwiązań dla strukturyzacji danych w branży budowlanej. W praktyce technologie te nie doprowadziły do powstania uniwersalnego rozwiązania, a jedynie dodały nowe komplikacje, a wysiłki te odzwierciedlają niesrealizowane ambicje koncepcji Semantycznego Internetu, gdzie oczekiwania znacznie przekroczyły rzeczywistość.

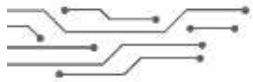


Rys. 6.1-10 Geometria i informacje w procesach budowlanych: od złożonych systemów CAD i BIM do uproszczonych danych do analizy.

Podczas gdy w branży IT niepowodzenia sieci semantycznej zostały zrekompensowane przez pojawienie się nowych technologii (big data, IoT, uczenie maszynowe, AR/VR), branża budowlana nie ma takich możliwości.

Oprócz wyzwań związanych z wykorzystaniem pojęć do komunikowania relacji danych między elementami projektu, pozostaje jeszcze jeden zasadniczy problem - sama dostępność tych danych. Branża budowlana jest nadal zdominowana przez zamknięte systemy, co utrudnia pracę z danymi, dzielenie się informacjami i poprawę wydajności procesów.

To właśnie zamknięty charakter danych staje się jedną z kluczowych barier utrudniających rozwój rozwiązań cyfrowych w budownictwie. W przeciwieństwie do branży IT, gdzie otwarte i zharmonizowane formaty danych stały się standardem, w sektorze CAD (BIM) każde oprogramowanie używa własnego formatu, tworząc zamknięte ekosystemy i sztucznie ograniczając użytkowników.



ROZDZIAŁ 5.2.

ZAMKNIĘTE FORMATY PROJEKTÓW I KWESTIE INTEROPERACYJNOŚCI

Zamknięte dane i spadająca produktywność: ślepy zaułek branży CAD (BIM)

Zastrzeżony charakter systemów CAD doprowadził do tego, że każdy program ma swój własny, unikalny format danych, który jest albo zamknięty i niedostępny z zewnątrz - RVT, PLN, DWG, NDW, NWD, SKP, lub jest dostępny w formie częściowo ustrukturyzowanej poprzez dość złożony proces konwersji - JSON, XML (CPIXML), IFC, STEP oraz ifcXML, IfcJSON, BIMJSON, IfcSQL, CSV, itd..

Różne formaty danych, w których można przechowywać te same dane o tych samych projektach, nie tylko różnią się strukturą, ale także zawierają różne wersje wewnętrznych znaczników, które programiści muszą wziąć pod uwagę, aby zapewnić kompatybilność aplikacji. Przykładowo, format CAD z 2025 roku otworzy się w programie CAD z 2026 roku, ale ten sam projekt nigdy nie otworzy się we wszystkich wersjach programu CAD, które mogły być dostępne przed 2025 rokiem.

Nie zapewniając bezpośredniego dostępu do baz danych, dostawca oprogramowania w branży budowlanej często tworzy własny unikalny format i narzędzia, których profesjonalista (inżynier projektu lub menedżer danych) musi użyć, aby uzyskać dostęp do danych, importować je i eksportować.

W konsekwencji sprzedawcy podstawowych systemów CAD (BIM) i powiązanych rozwiązań (np. ERP/PMIS)) stale podnoszą ceny za korzystanie z produktów, a zwykli użytkownicy są zmuszeni do płacenia "prowizji" na każdym etapie przesyłania danych według formatów [63]: za podłączanie, importowanie, eksportowanie i pracę z danymi, które użytkownicy sami stworzyli.

Koszt dostępu do danych w chmurze z popularnych produktów CAD - (BIM-) osiągnie 1 dolara za transakcję w 2025 roku [120], a subskrypcje na budowlane produkty ERP dla średnich firm osiągają pięcio- i sześciocyfrowe kwoty rocznie [121].

Istotą nowoczesnego oprogramowania budowlanego jest to, że to nie automatyzacja lub zwiększona wydajność, ale zdolność inżynierów do zrozumienia konkretnego wysoce wyspecjalizowanego oprogramowania, które wpływa na jakość i koszt przetwarzania danych projektu budowlanego, a także na zyski i długoterminowe przetrwanie firm podejmujących projekty budowlane.

Brak dostępu do baz danych CAD -systemów, które są wykorzystywane w dziesiątkach innych systemów i setkach procesów [63], a w konsekwencji brak wysokiej jakości komunikacji pomiędzy poszczególnymi specjalistami doprowadził branżę budowlaną do statusu jednego z najbardziej nieefektywnych sektorów gospodarki pod względem produktywności [44].

W ciągu ostatnich 20 lat aplikacji projektowych CAD- (BIM-), pojawienia się nowych systemów (ERP), nowych technologii budowlanych i materiałów, produktywność całego przemysłu budowlanego spadła o 20% (Rys. 2.2-1), podczas gdy ogólna produktywność wszystkich sektorów gospodarki, które nie mają większych problemów z dostępem do baz danych i marketingowych koncepcji BIM wzrosła o 70% (96% w

przemyśle wytwórczym) [122].



Rys. 6.2-1 Ze względu na izolację i złożoność danych projektowych, od których zależą dziesiątki działań i setki procesów w branży budowlanej, szybkość podejmowania decyzji jest kilkakrotnie wolniejsza niż w innych branżach.

Istnieją jednak również pojedyncze przykłady alternatywnych podejść do tworzenia interoperacyjności między rozwiązaniami CAD. Największa europejska firma budowlana z projektem SCOPE [123], rozpoczętym w 2018 roku, pokazuje, jak można wyjść poza klasyczną logikę systemów CAD- (BIM-). Zamiast próbować podporządkować IFC lub polegać na zastrzeżonych jądrach geometrii, programiści SCOPE używają interfejsów API i zestawów SDK inżynierii odwrotnej do ekstrakcji danych z różnych programów CAD, konwersji ich do neutralnych formatów, takich jak OBJ lub CPIXML w oparciu o jedyne jądro geometrii Open Source OCCT, a następnie stosują je w setkach procesów biznesowych firm budowlanych i projektowych. Jednak pomimo postępowości pomysłu, takie projekty napotykają na ograniczenia i złożoność wolnych jąder geometrycznych i nadal pozostają częścią zamkniętych ekosystemów jednej firmy, które odtwarzają logikę rozwiązań monovendor.

Ze względu na ograniczenia zamkniętych systemów i różnice w formatach danych, a także brak skutecznych narzędzi do ich ujednolicenia, firmy, które muszą pracować z formatami CAD, stają w obliczu gromadzenia znacznych ilości danych o różnym stopniu struktury i zamknięcia. Dane te nie są odpowiednio wykorzystywane i znikają w archiwach, gdzie pozostają na zawsze zapomniane i nieużywane.

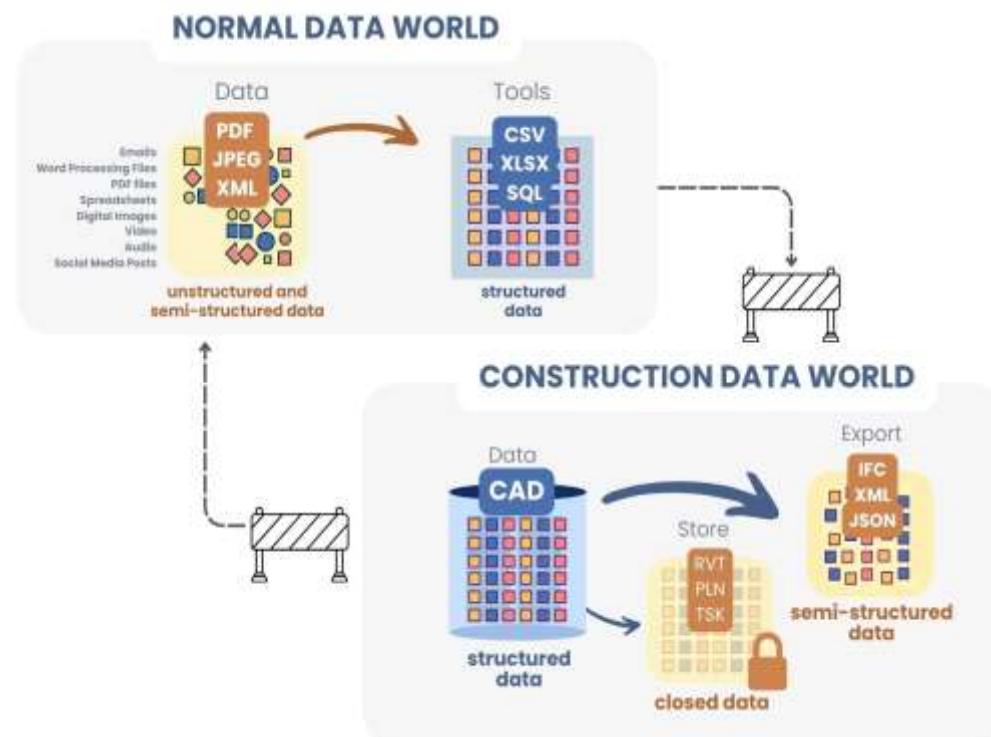
Dane uzyskane dzięki znacznemu wysiłkowi w fazie projektowania stają się niedostępne do dalszego wykorzystania ze względu na ich złożoność i zamknięty charakter.

W rezultacie, w ciągu ostatnich 30 lat, deweloperzy w branży budowlanej byli zmuszeni do ciągłego mierzenia się z tym samym problemem: każdy nowy zamknięty format lub zastrzeżone rozwiązanie generuje potrzebę integracji z istniejącymi otwartymi i zamkniętymi systemami CAD. Te ciągłe próby

zapewnienia interoperacyjności między różnymi rozwiązaniami CAD i BIM tylko komplikują ekosystem danych, zamiast przyczyniać się do jego uproszczenia i standaryzacji.

Mit interoperacyjności między systemami CAD

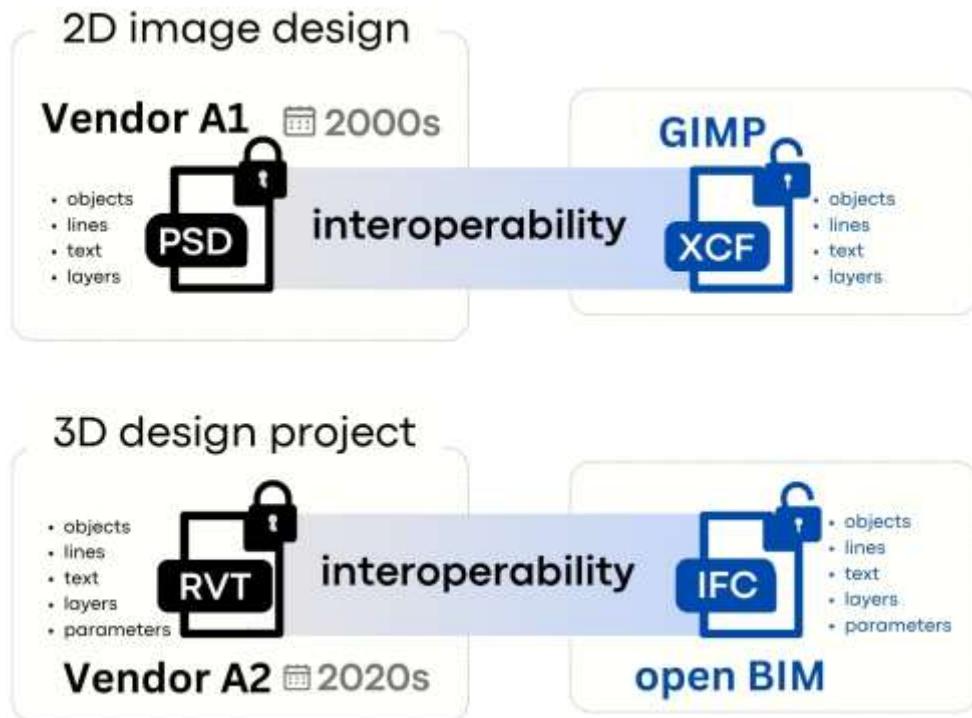
Podczas gdy w połowie lat 90. kluczowym kierunkiem rozwoju interoperacyjności w środowisku CAD było przełamanie zastrzeżonego formatu DWG - zakończone zwycięstwem sojuszu Open DWG [75] i faktycznym otwarciem najpopularniejszego formatu rysunkowego dla całej branży budowlanej - w połowie lat 2020. punkt ciężkości przesunął się. Nowy trend w branży budowlanej nabiera rozpędu: liczne zespoły programistów koncentrują się na tworzeniu tak zwanych "pomostów" między zamkniętymi systemami CAD (zamknięty BIM), formatem IFC i otwartymi rozwiązaniami (otwarty BIM). Większość z tych inicjatyw opiera się na wykorzystaniu formatu IFC i jądra geometrii OCCT, zapewniając techniczny pomost między różnymi platformami. Takie podejście jest postrzegane jako obiecujący kierunek, który może znacznie poprawić wymianę danych i interoperacyjność narzędzi programowych.



Rys. 6.2-2 Podczas gdy inne branże pracują z otwartymi danymi, branża budowlana musi pracować z zamkniętymi lub luźno ustrukturyzowanymi formatami CAD (BIM).

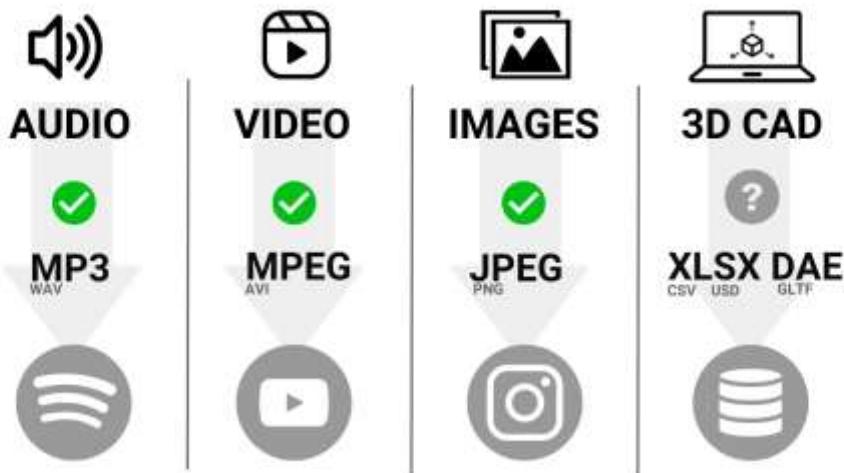
Takie podejście ma historyczne podobieństwa. W 2000 roku programiści, próbując przezwyciężyć dominację największego dostawcy edytorów graficznych (2D World), próbowali stworzyć płynną integrację między jego zastrzeżonym rozaniem a darmowym oprogramowaniem Open Source - alternatywą dla GIMP-a (rys. 6.2-3). W tamtych czasach, podobnie jak dziś w budownictwie, chodziło o próbę połączenia zamkniętych i otwartych systemów przy jednoczesnym zachowaniu złożonych parametrów, warstw i wewnętrznej logiki oprogramowania.

Jednak w rzeczywistości użytkownicy szukali prostych rozwiązań - płaskich, otwartych danych bez nadmiernej złożoności warstw i parametrów programu (analogi geometrycznego rdzenia w CAD). Użytkownicy poszukiwali prostych i otwartych formatów danych, wolnych od nadmiernej logiki. JPEG, PNG i GIF stały się takimi formatami w grafice. Dziś są one używane w sieciach społecznościowych, na stronach internetowych, w aplikacjach - są łatwe w przetwarzaniu i interpretacji, niezależnie od platformy lub dostawcy oprogramowania.



Rys. 6.2-3 Interoperacyjność formatów danych w budownictwie jest podobna do ścieżki prób połączenia zastrzeżonego produktu popularnego dostawcy i GIMP-a Open Source w 2000 roku.

W rezultacie prawie nikt w branży obrazowania nie używa dziś zamkniętych formatów, takich jak PSD lub otwarty XCF dla aplikacji, sieci społecznościowych, takich jak Facebook i Instagram, lub jako treści na stronach internetowych. Zamiast tego większość zadań wykorzystuje płaskie i otwarte formaty JPEG, PNG i GIF ze względu na łatwość użytkowania i szeroką kompatybilność. Otwarte formaty, takie jak JPEG i PNG, stały się standardem udostępniania obrazów ze względu na ich wszechstronność i szerokie wsparcie, dzięki czemu są łatwe w użyciu na różnych platformach. Podobną transformację można zaobserwować w przypadku innych formatów wymiany, takich jak wideo i audio, gdzie uniwersalne formaty, takie jak MPEG i MP3, są podkreślone ze względu na ich wydajność kompresji i szeroką kompatybilność. Taki ruch w kierunku standaryzacji uprościł udostępnianie i odtwarzanie treści i informacji, czyniąc je dostępnymi dla wszystkich użytkowników na wielu platformach (Rys. 6.2-4).



Rys. 6.2-4 Uproszczone formaty bez złożonych funkcji edycji stały się popularne w celu udostępniania i wykorzystywania danych.

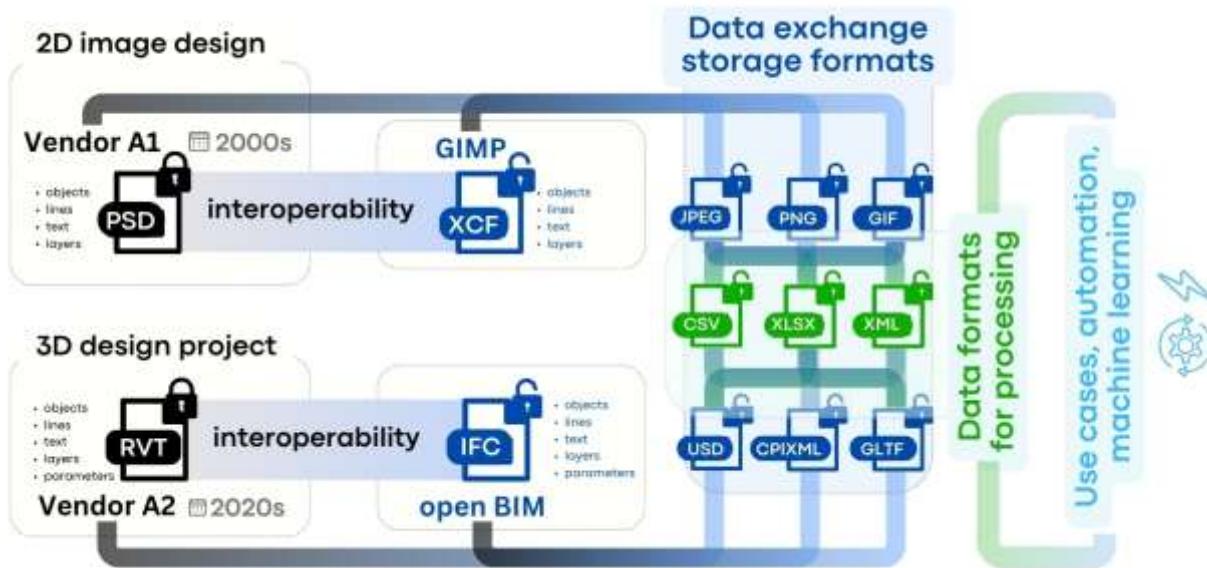
Podobne procesy zachodzą w modelowaniu 3D. Proste i otwarte formaty, takie jak USD, OBJ, glTF, DAE, DXF, SQL i XLSX są coraz częściej wykorzystywane w projektach do wymiany danych poza środowiskiem CAD (BIM). Formaty te przechowują wszystkie niezbędne informacje, w tym geometrię i metadane, bez konieczności obsługi złożonej struktury BREP, jądra geometrii lub wewnętrznych klasyfikatorów specyficznych dla dostawcy. Zastrzeżone formaty, takie jak NWC, SVF, SVF2, CPIXML i CP2 dostarczane przez wiodących producentów oprogramowania również pełnią podobne funkcje, ale pozostają zamknięte, w przeciwieństwie do otwartych standardów.

Warto zauważyć (i ponownie przypomnieć, jak już wspomniano w poprzednim rozdziale), że ten pomysł - odrzucenie pośrednich neutralnych i parametrycznych formatów, takich jak IGES, STEP i IFC - został poparty w 2000 roku przez głównego dostawcę CAD, który stworzył BIM Whitepaper i zarejestrował format IFC w 1994 roku. W Whitepaper 2000 "Zintegrowane projektowanie i produkcja" [65] dostawca CAD podkreśla znaczenie natywnego dostępu do bazy danych CAD w środowisku oprogramowania, bez konieczności korzystania z pośrednich translatorów i formatów parametrycznych, w celu zachowania kompletności i dokładności informacji.

Branża budowlana nie osiągnęła jeszcze porozumienia w sprawie narzędzi dostępu do baz danych CAD lub ich wymuszonej inżynierii odwrotnej, ani w sprawie przyjęcia wspólnego uproszczonego formatu danych do użytku poza platformami CAD (BIM). Na przykład wiele dużych firm w Europie Środkowej i regionach niemieckojęzycznych działających w sektorze budowlanym używa formatu CPIXML w swoich systemach ERP [121]. Ten zastrzeżony format, który jest rodzajem XML, łączy dane projektu CAD (BIM), w tym dane geometryczne i metadane, w jedną zorganizowaną, uproszoną strukturę. Duże firmy budowlane również tworzą własne formaty i systemy, jak w przypadku projektu SCOPE, który omówiliśmy w poprzednim rozdziale.

Zamknięta logika parametrycznych formatów CAD lub złożonych plików parametrycznych IFC (STEP) są zbędne w większości procesów biznesowych. Użytkownicy poszukują uproszczonych i płaskich formatów, takich jak USD, CPIXML, XML &OBJ, DXF, glTF, SQLite, DAE &XLSX, które zawierają wszystkie niezbędne

informacje o elementach, ale nie są obciążone zbędną logiką geometrii BREP, zależnością od jąder geometrii i wewnętrznymi klasyfikacjami określonych produktów CAD i BIM (rys. 6.2-5). 6.2-5).



Rys. 6.2-5 W większości przypadków użytkownicy wybierają najprostsze możliwe formaty, które są niezależne od programów dostawców.

Pojawienie się płaskich formatów obrazu, takich jak JPEG, PNG i GIF, wolnych od zbędnej logiki wewnętrznych silników producentów, napędziło rozwój tysięcy interoperacyjnych rozwiązań do przetwarzania i wykorzystywania grafiki. Doprowadziło to do pojawienia się aplikacji, począwszy od narzędzi do retuszowania i filtrowania, a skończywszy na sieciach społecznościowych, takich jak Instagram, Snapchat i Canva, w których te uproszczone dane mogą być wykorzystywane bez konieczności wiązania się z konkretnym twórcą oprogramowania.

Standaryzacja i uproszczenie formatów projektowych CAD będzie stymulować powstawanie wielu nowych, wygodnych i niezależnych narzędzi do pracy z projektami budowlanymi.

Odejście od złożonej logiki aplikacji dostawców powiązanych z zamkniętymi jądrami geometrii na rzecz uniwersalnych otwartych formatów opartych na bibliotekach uproszczonych elementów stwarza warunki wstępne dla bardziej elastycznej, przejrzystej i wydajnej obsługi danych. Otwiera to również dostęp do informacji dla wszystkich stron zaangażowanych w proces budowy - od projektantów po klientów i służby utrzymania ruchu.

Niemniej jednak jest wysoce prawdopodobne, że w nadchodzących latach dostawcy CAD będą próbować ponownie przenieść debatę na temat interoperacyjności i dostępu do baz danych CAD. Będzie ona już dotyczyć "nowych" koncepcji - takich jak dane granularne, inteligentne wykresy, "modele federacyjne", cyfrowe bliźniaki w repozytoriach chmurowych - a także tworzenia sojuszy branżowych i standardów, które kontynuują ścieżkę BIM i otwartego BIM. Pomimo atrakcyjnej terminologii, takie inicjatywy mogą ponownie stać się narzędziami do zatrzymywania użytkowników w ramach zastrzeżonych ekosystemów. Jednym z przykładów jest aktywna promocja formatu USD (Universal Scene Description) jako "nowego

standardu" wieloplatformowej współpracy CAD (BIM) od 2023 roku.

Przejdź do USD i danych granularnych m

Pojawienie się sojuszu AOUSD [124] w 2023 r. oznacza ważny zwrot w branży budowlanej. Jesteśmy świadkami początku nowej rzeczywistości, kształtowanej przez dostawców CAD, w zakresie postępowania z danymi budowlanymi poprzez kilka istotnych zmian. Pierwsza istotna zmiana dotyczy postrzegania danych CAD. Specjaliści zaangażowani we wczesne etapy projektowania koncepcyjnego coraz częściej zdają sobie sprawę, że tworzenie projektu w środowisku CAD to tylko punkt wyjścia. Dane wygenerowane podczas procesu projektowania ostatecznie stają się podstawą do analizy, obsługi i zarządzania obiekktami. Oznacza to, że muszą być one dostępne i użyteczne w systemach wykraczających poza tradycyjne narzędzia CAD.

Równolegle ma miejsce rewolucja w podejściu wiodących deweloperów. Wiodący w branży dostawca CAD, twórca koncepcji BIM i formatu IFC, dokonuje nieoczekiwanej zwrota w swojej strategii. Od 2023 r. firma odchodzi od tradycyjnego przechowywania danych w pojedynczych plikach, koncentrując się na pracy z danymi granularnymi (znormalizowanymi i ustrukturyzowanymi) i przechodząc na podejście skoncentrowane na danych [125].

Sprzedawcy podążają za historycznymi trendami innych branż: większość użytkowników nie potrzebuje zamkniętych formatów CAD (podobnych do PSD) lub złożonych parametrycznych plików IFC (podobnych do GIMP z logiką warstw). Potrzebują prostych obrazów obiektów, które można wykorzystać w CAFM (Instagram budowlany), ERP (Facebook) i tysiącach innych procesów wypełnionych arkuszami kalkulacyjnymi Excel i dokumentami PDF.

Obecne trendy w branży budowlanej potencjalnie przygotowują grunt pod stopniowe odchodzenie od parametrycznych i złożonych formatów na rzecz bardziej uniwersalnych i niezależnych formatów USD, GLTF, DAE, OBJ (z metainformacjami zarówno w ramach hybrydy, jak i w oddzielnych formatach strukturalnych lub luźno ustrukturyzowanych). Historyczni liderzy, w tym główne firmy projektowe, które kiedyś aktywnie promowały IFC w połowie lat 90-tych, obecnie otwarcie promują nowy format USD [93], podkreślając jego prostotę i wszechstronność (rys. 6.2-6). Masowe przyjęcie USD w produktach, kompatybilność z GLTF i aktywna integracja z narzędziami takimi jak Blender, Unreal Engine i Omniverse pokazują potencjał początku nowego paradymatu pracy z danymi. Wraz z popularnością lokalnych rozwiązań, takich jak europejski płaski format USD - CPIXML, używany w popularnych europejskich systemach ERP może potencjalnie wzmacnić pozycję USD w Europie Środkowej. Organizacje zaangażowane w rozwój formatu IFC już dostosowują swoją strategię do USD [126], co tylko potwierdza nieuchronność tej zmiany.

Technical Specifications				Comparison / Notes
File Structure	Monolithic file	Uses ECS and linked data	IFC stores all data in one file; USD uses Entity-Component-System and linked data for modularity and flexibility	
Data Structure	Complex semantics, parametric geometry	Flat format, geometry in MESH, data in JSON	IFC is complex and parametric; USD is simpler and uses flat data	
Geometry	Parametric, dependent on BREP	Flat, MESH (triangular meshes)	IFC uses parametrics; USD uses meshes for simplified processing.	
Properties	Complex structure of semantic descriptions	Properties in JSON, easy access	Properties in USD are easier to use thanks to JSON	
Export/import	Complex implementation, dependent on third-party SDKs	Easy integration, wide support	USD integrates more easily and is supported in many products	
Format Complexity	High, requires deep understanding	Low, optimized for convenience	The time required to understand the structure of the file and the information stored in it.	
Performance	Can be slow when processing large models	High performance in visualization and processing	USD is optimized for speed and efficiency. Simulations, machine learning, AI, smart cities will be held in the Nvidia Omniverse	
Integration with 3D Engines	Limited	High, designed for graphics engines	USD excels with native support for real-time visualization platforms	
Support outside CAD Software	BlenderBIM, IfcOpenShell	Unreal Engine, Unity, Blender, Omniverse	USD is widely supported in graphics tools	
Cloud Technology Support	Limited	Well-suited for cloud services and online collaboration	USD is optimized for cloud solutions	
Ease of Integration into Web Applications	Difficult to integrate due to size and complexity	Easy to integrate, supports modern web technologies	USD is preferable for web applications	
Change Management	Versions through separate files	Versioning built into the format core	IFC handles changes via separate files, while USD embeds versioning directly into its structure	
Collaboration Support	Supports data exchange between project participants	Designed for collaborative work on complex scenes	USD provides efficient collaboration through layers and variations	
Learnability	Steep learning curve due to complexity	Easier to master thanks to a clear structure	USD is easier to learn and implement	

Rys. 6.2-6 Porównanie specyfikacji technicznych formatów IFC i USD.

W tym kontekście USD ma potencjał, aby stać się de facto standardem, obiecując przewyściąć wiele obecnych ograniczeń, związanych przede wszystkim ze złożonością istniejących formatów CAD - (BIM-) i zależnością ich interpretacji od jąder geometrycznych.

Zamiast parametrycznych i złożonych formatów CAD i IFC - uproszczone formaty danych USD, glTF, DAE, OBJ z metainformacjami o elementach w CSV, XLSX, JSON, XML zyskają miejsce w branży budowlanej ze względu na ich prostotę i elastyczność.

Obecne zmiany w branży budowlanej na pierwszy rzut oka wyglądają jak przełom technologiczny związany z przejściem z przestarzałego IFC na nowocześniejszy USD. Warto jednak wziąć pod uwagę, że jeszcze w 2000 roku ten sam producent CAD, który opracował IFC, pisał o jego problemach i potrzebie dostępu do

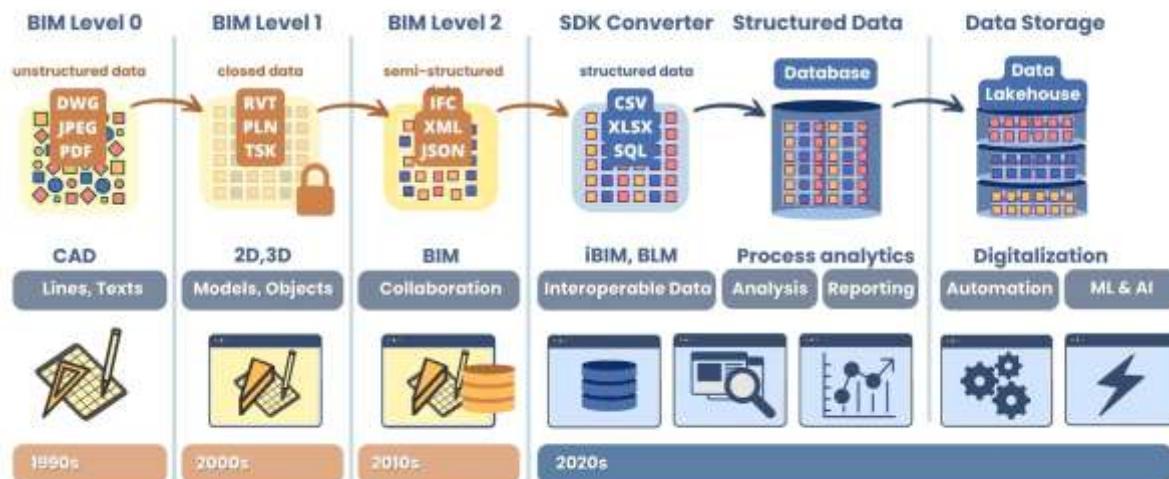
bazy danych [65], a obecnie aktywnie promuje przejście na nowy standard - USD.

Za kolejną fasadą "otwartych danych" USD i "nowych" koncepcji zarządzania danymi granularnymi, za pośrednictwem aplikacji opartych na chmurze, które dostawcy CAD zaczynają promować, może kryć się zamiar dostawców zmonopolizowania zarządzania danymi projektowymi, gdzie użytkownicy znajdują się w sytuacji, w której wybór formatu jest bardziej związany z interesami korporacyjnymi niż z rzeczywistymi potrzebami.

Analiza kluczowych faktów [93] pokazuje, że głównym celem tych zmian jest raczej mniejsza wygoda użytkownika niż przede wszystkim utrzymanie kontroli nad ekosystemami i przepływami danych na korzyść dostawców, którzy przez 40 lat nigdy nie byli w stanie zapewnić dostępu do baz danych CAD.

Być może nadszedł czas, aby firmy przestały czekać na nowe koncepcje od dostawców oprogramowania i skupiły się na samodzielnym rozwoju w kierunku skoncentrowanym na danych. Uwolniwszy się od problemów z dostępem do danych dzięki narzędziom inżynierii wstępnej, branża będzie mogła samodzielnie przejść do nowoczesnych, darmowych i wygodnych narzędzi do pracy i analizy danych bez narzucania nowych koncepcji.

CAD (BIM) Maturity Levels: From Stage 0 to Structured Data



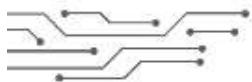
Rys. 6.2-7 Poziom dojrzałości CAD (BIM): od danych niestrukturyzowanych do danych ustrukturyzowanych i repozytoriów.

Dostęp do baz danych, otwartych danych i formatów nieuchronnie stanie się standardem w branży budowlanej, niezależnie od prób zatrzymania tego procesu przez sprzedawców - to tylko kwestia czasu (rys. 6.2-7). Tempo tego przejścia może znacznie wzrosnąć, jeśli coraz więcej specjalistów zapozna się z otwartymi formatami, narzędziami bazodanowymi i dostępnymi zestawami SDK do inżynierii odwrotnej, umożliwiającymi bezpośredni dostęp do danych CAD -systemy [92].

Przyszłość leży w otwartych, ujednoliconych i analitycznie dostępnych danych. Aby uniknąć uzależnienia od rozwiązań dostawców i nie stać się zakładnikiem zamkniętych ekosystemów, firmy budowlane i inżynierijne przdzej czy później będą musiały postawić na otwartość i niezależność, wybierając formaty i rozwiązania zapewniające pełną kontrolę nad danymi.

Dane, które powstają dziś w branży budowlanej, będą w przyszłości kluczowym źródłem decyzji biznesowych. Będą stanowić strategiczne "paliwo" napędzające rozwój i wydajność firm budowlanych. Przyszłość branży budowlanej leży w umiejętności pracy z danymi, a nie w wyborze formatów lub modeli danych.

Aby zrozumieć różnicę między otwartymi formatami USD, glTF, DAE, OBJ i zastrzeżonymi formatami parametrycznymi CAD, ważne jest, aby wziąć pod uwagę jeden z najbardziej złożonych i kluczowych elementów danych w wizualizacji i obliczeniach projektowych - geometrię i procesy jej generowania. Aby zrozumieć, w jaki sposób dane geometryczne stają się podstawą analityki i obliczeń w budownictwie, konieczne jest głębsze przyjrzenie się mechanizmom generowania, przekształcania i przechowywania geometrii.



ROZDZIAŁ 5.3.

GEOMETRIA W BUDOWNICTWIE: OD LINII DO METRÓW SZEŚCIENNYCH

Kiedy linie zamieniają się w pieniądze lub dlaczego budowniczowie potrzebują geometrii

Geometria w budownictwie to nie tylko wizualizacja, ale także podstawa dokładnych obliczeń ilościowych. W modelu projektu geometria uzupełnia listy parametrów elementów (rys. 3.1-16) o ważne charakterystyki objętościowe, takie jak długość, powierzchnia i objętość. Te wartości parametrów objętościowych są obliczane automatycznie przy użyciu jąder geometrii i stanowią punkt wyjścia dla szacunków, harmonogramów i modeli zasobów. Jak już omówiliśmy w części 5 tej książki oraz w rozdziale "Kosztorysowanie i szacowanie projektów budowlanych", to właśnie parametry objętościowe grup obiektów z modeli CAD stanowią podstawę nowoczesnych systemów ERP, PMIS Geometria odgrywa fundamentalną rolę nie tylko w fazie projektowania, ale także w zarządzaniu realizacją projektu, kontroli harmonogramu, budżetowaniu i eksploatacji. Tak jak tysiące lat temu, podczas budowy egipskich piramid, dokładność projektu zależała od miar długości, takich jak łokcie i kubity, tak dziś dokładność interpretacji geometrii w programach CAD bezpośrednio wpływa na wynik: od budżetu i terminów po wybór wykonawcy i logistykę dostaw

W bardzo konkurencyjnym i ograniczonym budżetowo środowisku, dokładność obliczeń objętościowych, która bezpośrednio zależy od geometrii, staje się czynnikiem przetrwania. Nowoczesne systemy ERP - bezpośrednio zależą od prawidłowych charakterystyk objętościowych uzyskanych z modeli CAD - i BIM -. Dlatego dokładny opis geometryczny elementów nie jest tylko wizualizacją, ale kluczowym narzędziem do zarządzania kosztami i czasem budowy.

W przeszłości geometria była podstawowym językiem komunikacji inżynierskiej. Od linii na papirusie po modele cyfrowe, rysunki i reprezentacje geometryczne służyły jako środek wymiany informacji między projektantami, brygadistami i kosztorysantami. Przed pojawiением się komputerów obliczenia były wykonywane ręcznie, przy użyciu linijk i kątomierzy. Dziś zadanie to jest zautomatyzowane dzięki modelowaniu wolumetrycznemu: geometryczne jądra oprogramowania CAD przekształcają linie i punkty w trójwymiarowe bryły, z których automatycznie wyodrębniane są wszystkie niezbędne cechy.

Pracując w programach CAD, tworzenie elementów geometrycznych do obliczeń odbywa się za pośrednictwem interfejsu użytkownika programów CAD- (BIM-). Do przekształcania punktów i linii w bryły wolumetryczne wykorzystywane jest jądro geometryczne, które spełnia kluczowe zadanie - przekształcanie geometrii w modele wolumetryczne, na podstawie których po aproksymacji automatycznie obliczana jest charakterystyka wolumetryczna elementu.

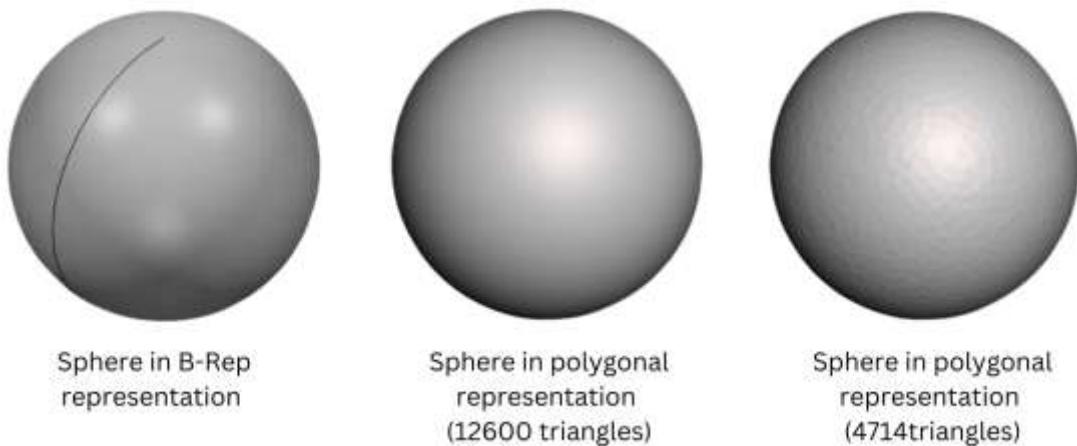
Od linii do objętości: Jak powierzchnia i objętość stają się danymi

W praktyce inżynierskiej objętości i obszary są obliczane na podstawie powierzchni geometrycznych opisanych analitycznie lub za pomocą modeli parametrycznych, takich jak NURBS (niejednolite racjonalne

B-splajny) w ramach BREP (reprezentacja elementów granicznych).

NURBS (Non-Uniform Rational B-Splines) to matematyczny sposób opisywania krzywych i powierzchni, podczas gdy BREP to struktura do opisywania pełnej trójwymiarowej geometrii obiektu, w tym jego granic, które można zdefiniować za pomocą NURBS.

Pomimo dokładności BREP i NURBS, wymagają one potężnych zasobów obliczeniowych i złożonych algorytmów. Jednak bezpośrednie obliczenia na podstawie takich matematycznie dokładnych opisów są często trudne obliczeniowo, więc w praktyce prawie zawsze stosuje się teselację - przekształcenie powierzchni w siatkę trójkątów - co upraszcza późniejsze obliczenia. Teselacja to podział złożonej powierzchni na trójkąty lub wielokąty. W środowiskach CAD /CAE metoda ta jest używana do wizualizacji, obliczeń objętości, wyszukiwania kolizji, eksportu do formatów takich jak MESH i analizy kolizji. Przykładem z natury jest plaster miodu pszczole, w którym złożony kształt jest podzielony na regularną siatkę (rys. 6.3-1).



Rys. 6.3-1 Ta sama kula w opisie parametrycznym BREP i reprezentacji wielokątnej z różną liczbą trójkątów.

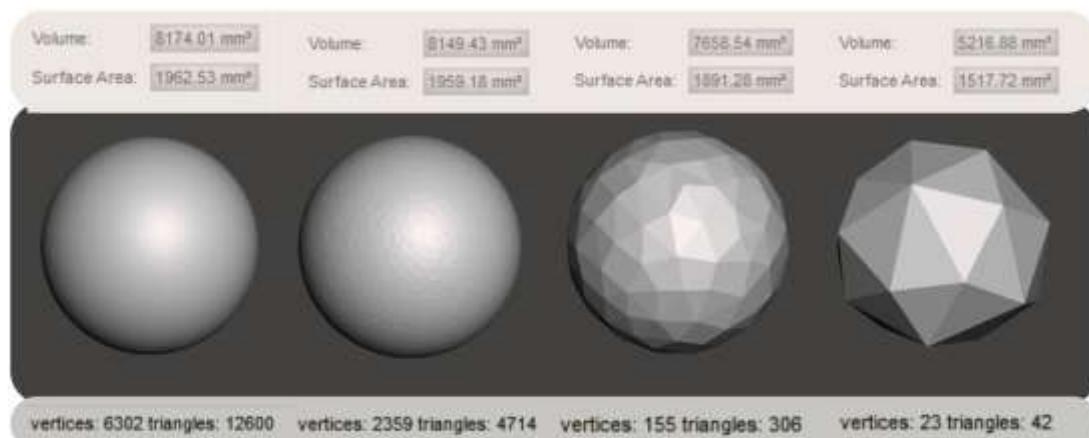
BREP (NURBS), używany w CAD, nie jest podstawowym modelem geometrii. Został stworzony jako wygodne narzędzie do reprezentowania okręgów i racjonalnych splajnów oraz do minimalizacji przechowywania danych geometrycznych. Ma jednak ograniczenia - na przykład niezdolność do dokładnego opisu sinusoidy, która leży u podstaw linii śrubowych i powierzchni, oraz potrzebę stosowania złożonych jąder geometrycznych.

Z kolei siatki trójkątne i teselacja kształtów parametrycznych charakteryzują się prostotą, efektywnym wykorzystaniem pamięci i możliwością przetwarzania dużych ilości danych (rys. 6.3-2). Zalety te sprawiają, że podczas obliczania kształtów geometrycznych można obejść się bez skomplikowanych i kosztownych jąder geometrycznych oraz dziesiątek milionów linii kodu w nich osadzonych.

W większości przypadków budynków nie ma znaczenia, jak dokładnie zdefiniowana jest charakterystyka objętościowa - za pomocą modeli parametrycznych (BREP, IFC) lub za pomocą wielokątów (USD, glTF, DAE, OBJ). Geometria pozostaje formą przybliżenia: czy to poprzez NURBS czy MESH, zawsze jest to przybliżony opis kształtu.

Geometria zdefiniowana jako wielokąty lub BREP (NURBS) pozostaje do pewnego stopnia jedynie sposobem przybliżenia za pomocą przybliżonego opisu ciągłego kształtu. Tak jak całki Fresnela nie mają dokładnego wyrażenia analitycznego, tak dyskretyzacja geometrii za pomocą wielokątów lub NURBS jest zawsze przybliżeniem, podobnie jak trójkątne MESH.

Geometria parametryczna w formacie BREP jest niezbędna głównie tam, gdzie ważny jest minimalny rozmiar danych i możliwe jest wykorzystanie zasobochłonnych i drogich jąder geometrii do ich przetwarzania i wyświetlania. Najczęściej jest to charakterystyczne dla twórców programów CAD, którzy w tym celu stosują w swoich produktach jądra geometryczne dostawców MCAD. W tym przypadku, nawet w tych programach, modele BREP w procesie teselacji do wizualizacji i obliczeń są często konwertowane na trójkąty (podobnie jak pliki PSD są upraszczane do JPEG).



Rys. 6.3-2 Różnica charakterystyk objętościowych w figurach o różnej liczbie wielokątów.

Wieloboczny MESH, jak i parametryczny BREP, mają swoje zalety i ograniczenia, ale cel jest ten sam - opisać geometrię z uwzględnieniem zadań użytkownika. Ostatecznie dokładność modelu geometrycznego zależy nie tylko od metody jego reprezentacji, ale także od wymagań konkretnego zadania.

W większości problemów konstrukcyjnych, potrzeba geometrii parametrycznej i złożonych jąder geometrycznych może być zbędna.

W każdym konkretnym zadaniu automatyzacji obliczeń warto zastanowić się, czy znaczenie geometrii parametrycznej nie jest wyolbrzymiane przez programistów CAD, którzy są zainteresowani promowaniem i sprzedażą własnego oprogramowania.

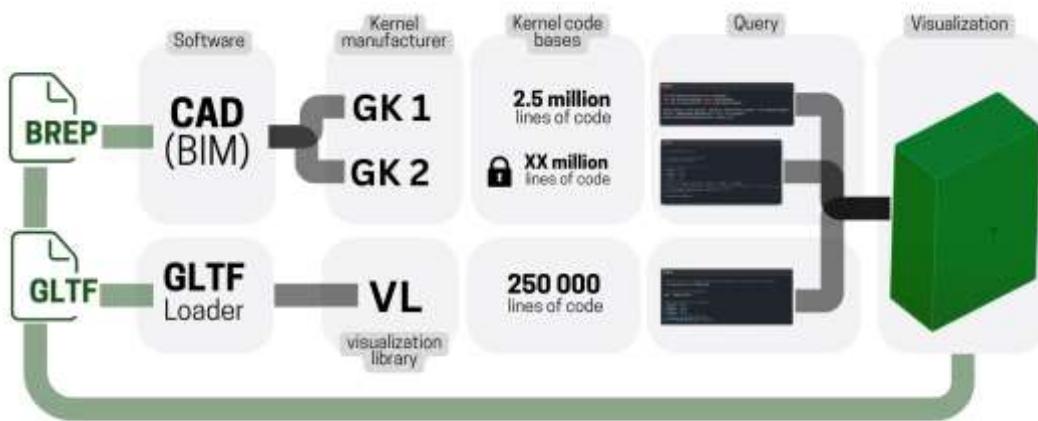
Przejście na MESH, USD i wielokąty: korzystanie z teselacji dla geometrii

W branży budowlanej, podczas przesyłania strumieniowego, tworzenia systemów, baz danych lub automatyzacji procesów, które pracują z informacjami projektowymi i geometrią elementów, ważne jest, aby dążyć do niezależności od konkretnych edytorów CAD i jąder geometrii.

Format wymiany stosowany zarówno w działach obliczeniowych, jak i na placu budowy nie powinien być oparty na konkretnym programie CAD- (BIM-). Informacje geometryczne powinny być reprezentowane w formacie bezpośrednio poprzez teselację, bez odniesienia do rdzenia geometrycznego lub architektury CAD.

Geometria parametryczna z CAD może być traktowana jako źródło pośrednie, ale nie jako podstawa uniwersalnego formatu. Większość opisów parametrycznych (w tym BREP i NURBS) jest w każdym przypadku konwertowana na wielokąty MESH w celu dalszego przetwarzania. Jeśli wynik jest taki sam (teselacja i wielokąty), a proces jest prostszy, wybór jest oczywisty. Jest to podobne do wyboru między ontologiami grafów a tabelami strukturalnymi (które omówiliśmy w części czwartej): nadmierna złożoność rzadko jest uzasadniona (rys. 3.2-10, rys. 6.1-8).

Otwarte formaty, takie jak: OBJ, STL, glTF, SVF, CPIXML, USD i DAE, wykorzystują uniwersalną strukturę siatki trójkątów, co daje im znaczące korzyści. Formaty te charakteryzują się doskonałą interoperacyjnością - są łatwe do odczytu i wizualizacji przy użyciu dostępnych bibliotek open source bez potrzeby stosowania złożonych, wyspecjalizowanych jąder geometrii zawierających miliony linii kodu (Rys. 6.3-3). Te wszechstronne formaty geometrii są wykorzystywane w zastosowaniach od stosunkowo prostych narzędzi do projektowania kuchni w IKEA™ po złożone systemy wizualizacji obiektów w kinie i aplikacjach VR. Ważną zaletą jest dostępność dużej liczby darmowych i otwartych bibliotek do pracy z tymi formatami, dostępnych dla większości platform i języków programowania.



Rys. 6.3-3 Tą samą reprezentację geometrii uzyskuje się poprzez zastosowanie formatów parametrycznych i jąder geometrii lub poprzez zastosowanie formatów triangulacyjnych i bibliotek wizualizacji open source.

Podobnie jak sami użytkownicy, dostawcy CAD napotykają problemy z interpretacją obcych parametrycznych formatów CAD lub otwartego IFC z powodu różnych jąder geometrycznych. W praktyce

wszyscy dostawcy CAD bez wyjątku używają zestawu SDK do inżynierii odwrotnej do przesyłania danych między systemami i żaden z nich nie polega na formatach takich jak IFC lub USD [93] do celów interoperacyjności.

Zamiast korzystać z koncepcji promowanych przez sojusze dostawców CAD, z których sami nie korzystają - bardziej produktywne dla programistów i użytkowników rozwiązań CAD jest skupienie się na zrozumieniu korzyści płynących z każdego podejścia w określonym kontekście i wybranie jednego lub drugiego rodzaju geometrii w zależności od przypadku użycia. Wybór między różnymi reprezentacjami geometrycznymi jest kompromisem między dokładnością, wydajnością obliczeniową i praktycznymi potrzebami konkretnego zadania.

Złożoność związana z wykorzystaniem geometrycznych jąder, tradycyjnie narzucanych branży budowlanej przez dużych dostawców podczas przetwarzania danych projektowych, często okazuje się zbędna. Format USD oparty na geometrii MESH może stać się swego rodzaju "puszką Pandory" dla branży, otwierając przed programistami nowe możliwości organizacji wymiany danych - poza ramami IFC i parametrycznych struktur BREP typowych dla dostawców CAD.

Po bliższym przyjrzeniu się strukturze USD, DAE, glTF, OBJ itp. staje się oczywiste, że istnieją prostsze, otwarte formaty, które pozwalają efektywnie organizować transfer i wykorzystanie informacji geometrycznych bez konieczności polegania na złożonej parametryce i zamkniętych jądrach geometrycznych. Takie podejście nie tylko obniża techniczny próg wejścia dla deweloperów, ale także sprzyja rozwojowi elastycznych, skalowalnych i prawdziwie otwartych rozwiązań dla budownictwa cyfrowego.

LOD, LOI, LOMD - unikalna klasyfikacja detali w CAD (BIM)

Oprócz formatów reprezentacji geometrycznej, w świecie, w którym różne branże wykorzystują różne poziomy szczegółowości i głębokości danych, metodologie CAD - (BIM-) oferują własne unikalne systemy klasyfikacji, które strukturyzują podejście do informowania modeli budynków.

Jednym z przykładów nowego podejścia do standaryzacji jest wprowadzenie poziomów rozwoju modelu, odzwierciedlających stopień gotowości i wiarygodności zarówno komponentów graficznych, jak i informacyjnych. Dla rozróżnienia zawartości informacyjnej w pracy z danymi CAD - (BIM-) pojawił się LOD (Level Of Detail) - poziom szczegółowości części graficznej modelu, oraz LOI (Level Of Information) - poziom opracowania danych. Dodatkowo dla podejścia zintegrowanego wprowadzono pojęcie LOA (Level of Accuracy) - dokładność prezentowanych elementów oraz LOG (Level of Geometry) do określenia dokładności reprezentacji graficznej.

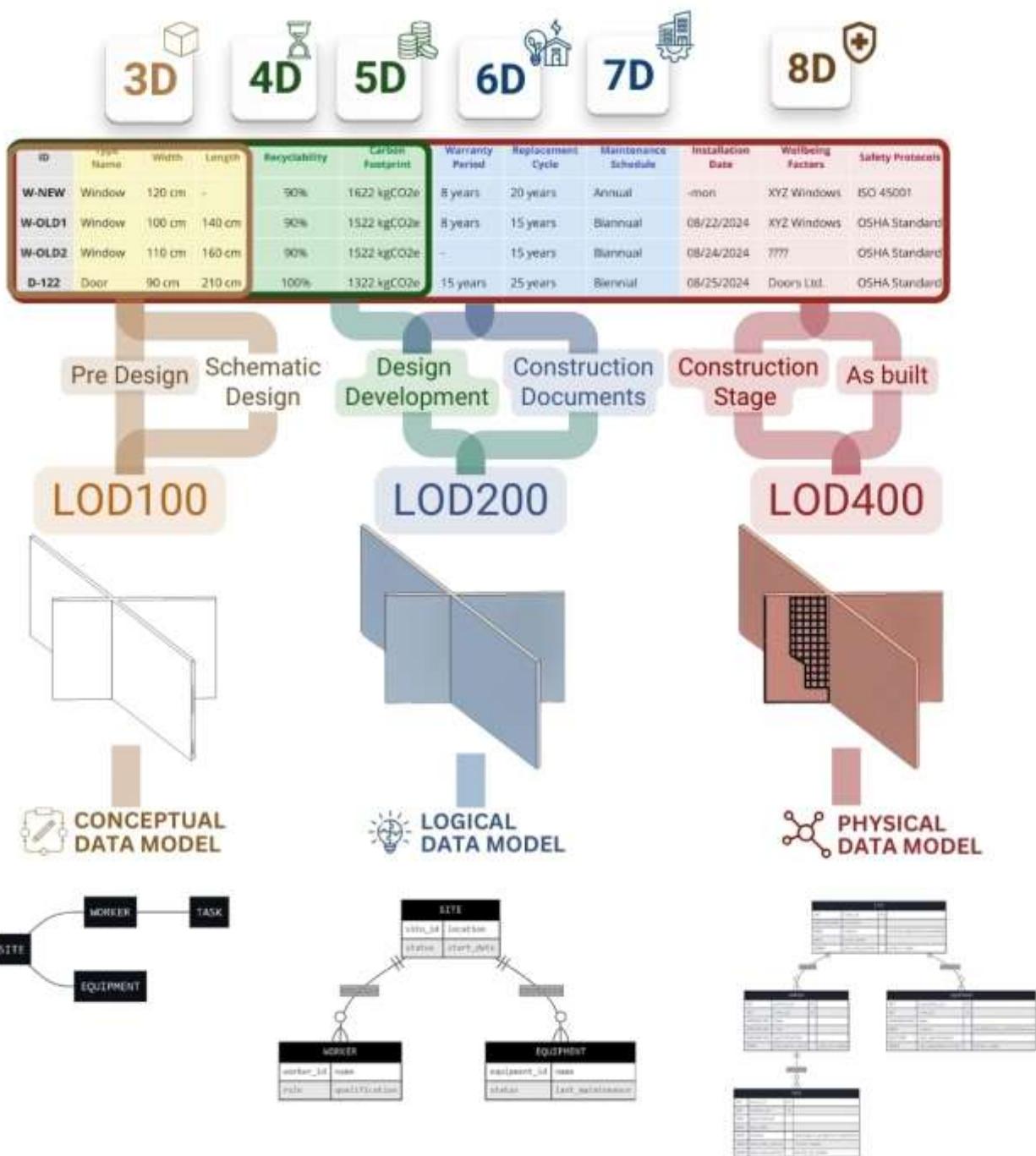
Poziomy szczegółowości (LOD) są oznaczone liczbami od 100 do 500, odzwierciedlającymi stopień rozwoju modelu. LOD 100 to model koncepcyjny z ogólnymi kształtami i wymiarami. LOD 200 zawiera bardziej precyzyjne wymiary i kształty, ale z warunkową szczegółowością. LOD 300 to szczegółowy model z dokładnymi wymiarami, kształtami i lokalizacjami elementów. LOD 400 zawiera szczegółowe informacje niezbędne do produkcji i instalacji elementów. LOD 500 odzwierciedla rzeczywisty stan obiektu po zakończeniu budowy i jest używany do obsługi i konserwacji. Poziomy te opisują strukturę nasycenia informacjami modelu CAD (BIM) na różnych etapach cyklu życia, w tym 3D, 4D, 5D i dalej.

W rzeczywistych projektach wysoki poziom szczegółowości (LOD400) jest często nadmierny i wystarczy

użyć geometrii LOD100 lub nawet płaskich rysunków, podczas gdy resztę danych można uzyskać obliczeniowo lub z powiązanych elementów, które mogą nie mieć wyraźnej geometrii. Na przykład przestrzenie i elementy pomieszczeń (kategorie elementów pomieszczeń) mogą nie mieć geometrii wizualnej, ale zawierają znaczne ilości informacji i baz danych, wokół których zbudowanych jest wiele procesów biznesowych.

Dlatego ważne jest, aby jasno zdefiniować wymagany poziom szczegółowości przed rozpoczęciem projektowania. Dla przypadków użycia 4D -7D często wystarczające są nawet rysunki DWG i minimalna geometria LOD100. Kluczowym zadaniem w procesie tworzenia wymagań jest znalezienie równowagi pomiędzy bogactwem i praktycznością modelu.

W istocie, jeśli uznamy dane CAD (BIM) za bazę danych (która są), to opis nasycenia modelu nowymi akronimami jest niczym innym, jak modelowaniem danych systemów informatycznych krok po kroku, począwszy od poziomu koncepcyjnego, a skończywszy na fizycznym (rys. 6.3-4), co zostało szczegółowo omówione w trzeciej i czwartej części książki. Każdy wzrost LOD i LOI oznacza dodanie informacji potrzebnych do nowych zadań: obliczeń, zarządzania budową, eksploatacji i charakteryzuje się sukcesywnym wzbogacaniem modelu o dodatkowe warstwy informacyjne (3D -8D) w postaci różnych parametrów, które omówiliśmy w piątej części książki.



Rys. 6.3-4 Proces dopracowywania szczegółów projektu jest identyczny z modelowaniem danych od koncepcyjnego do fizycznego modelu danych.

Geometria to tylko część danych projektowych, których potrzeba nie zawsze jest uzasadniona w projektach budowlanych, a kluczową kwestią pracy z danymi CAD jest nie tyle sposób wizualizacji modeli, ale bardziej sposób, w jaki dane z tych modeli mogą być wykorzystywane poza programami CAD- (BIM -).

W połowie XXI wieku branża budowlana stanęła przed bezprecedensowym wyzwaniem związanym z gwałtownym wzrostem ilości danych w systemach zarządzania i przetwarzania danych, zwłaszcza tych

pochodzących z działów CAD (BIM). Ten dramatyczny wzrost ilości danych zaskoczył menedżerów firm i nie byli oni przygotowani na rosnące wymagania dotyczące jakości danych i zarządzania nimi.

Nowe standardy CAD (BIM) - AIA, BEP, IDS, LOD, COBie

Korzystając z braku otwartego dostępu do baz danych CAD i ograniczonej konkurencji na rynku przetwarzania danych, a także wykorzystując kampanie marketingowe związane z nowym akronimem BIM, organizacje zaangażowane w rozwój podejścia do postępowania z danymi CAD - zaczęły tworzyć nowe standardy i koncepcje, które de iure powinny mieć na celu poprawę praktyk zarządzania danymi.

Chociaż prawie wszystkie inicjatywy bezpośrednio lub pośrednio wspierane przez dostawców CAD i deweloperów (BIM) miały na celu optymalizację przepływu pracy, zaowocowały one mnóstwem standardów lobbowanych przez różnych interesariuszy, co doprowadziło branżę budowlaną do pewnej niejednoznaczności i zamieszania w zakresie procesów danych.

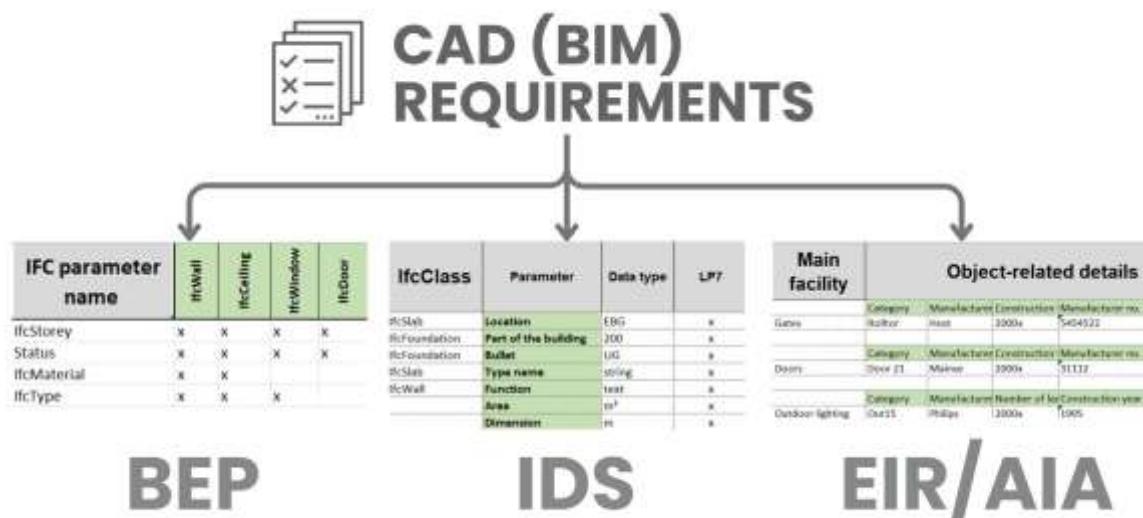
Wymieńmy niektóre z nowych standardów danych, oprócz LOD, LOI, LOA, LOG, które pojawiły się w ostatnich latach w branży budowlanej:

- **BEP** (BIM Execution Plan) - opisuje sposób integracji i wykorzystania CAD (BIM) w projekcie, definiując metody i procesy obsługi danych.
- **Dokument EIR /AIA** (Wymagania informacyjne klienta) - przygotowany przez klienta przed zaproszeniem do składania ofert i zawierający wymagania dla wykonawcy w zakresie przygotowania i dostarczenia informacji. Służy on jako podstawa dla BEP w danym projekcie.
- **AIM** (Asset Information Model) jest częścią procesu BIM. Po dostarczeniu i ukończeniu projektu model danych nazywany jest modelem informacji o aktywach lub AIM. Celem AIM jest zarządzanie, utrzymanie i obsługa zrealizowanego zasobu.
- **IDS** (Information Delivery Specification) - definiuje wymagania oraz jakie dane i w jakim formacie są wymagane na różnych etapach projektu budowlanego.
- **iLOD** to poziom szczegółowości LOD, z jakim informacje są reprezentowane w modelu BIM. Określa, jak szczegółowe i kompletne są informacje w modelu, od podstawowych reprezentacji geometrycznych po szczegółowe specyfikacje i dane.
- **eLOD** - LOD poziom szczegółowości poszczególnych elementów w modelu CAD (BIM). Określa zakres, w jakim każdy element jest modelowany i powiązane informacje, takie jak wymiary, materiały, charakterystyka wydajności i inne istotne atrybuty.
- **APS** (Platform Services) i inne produkty głównych dostawców CAD (BIM) - opisują narzędzia i infrastrukturę potrzebną do tworzenia połączonych i otwartych modeli danych.

Chociaż deklarowanym celem wdrażania standardów CAD (BIM) - takich jak LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD - jest poprawa jakości zarządzania danymi i rozszerzenie możliwości automatyzacji, w praktyce ich stosowanie często prowadzi do nadmiernej złożoności i fragmentacji procesów. Jeśli potraktujemy model CAD (BIM) jako rodzaj bazy danych, staje się oczywiste, że wiele z tych standardów powiela od dawna znane i skuteczne podejścia stosowane w innych branżach w pracy z systemami informatycznymi. Zamiast uproszczenia i ujednolicenia, takie inicjatywy często tworzą

dodatkowe obciążenie terminologiczne i utrudniają wdrażanie prawdziwie otwartych i elastycznych rozwiązań.

Warto zauważyc, że wiele z tych nowych koncepcji w rzeczywistości zastępuje procesy modelowania i walidacji danych, które zostały szczegółowo omówione w pierwszych częściach książki i które od dawna są stosowane w innych sektorach gospodarki. Z drugiej strony, w budownictwie proces standaryzacji często zmierza w przeciwnym kierunku - tworzone są nowe formaty opisu danych, nowe standardy i nowe koncepcje walidacji danych, które nie zawsze prowadzą do rzeczywistej jednolitości i praktycznego zastosowania. W rezultacie, zamiast uproszczenia i automatyzacji przetwarzania, branża staje w obliczu dodatkowych poziomów regulacji i biurokracji (rys. 6.3-1), co nie zawsze sprzyja zwiększeniu wydajności.



Rys. 6.3-1 Wymagania dotyczące zawartości danych i informacji są zredukowane do opisu atrybutów i ich wartości granicznych, opisanych za pomocą tabel.

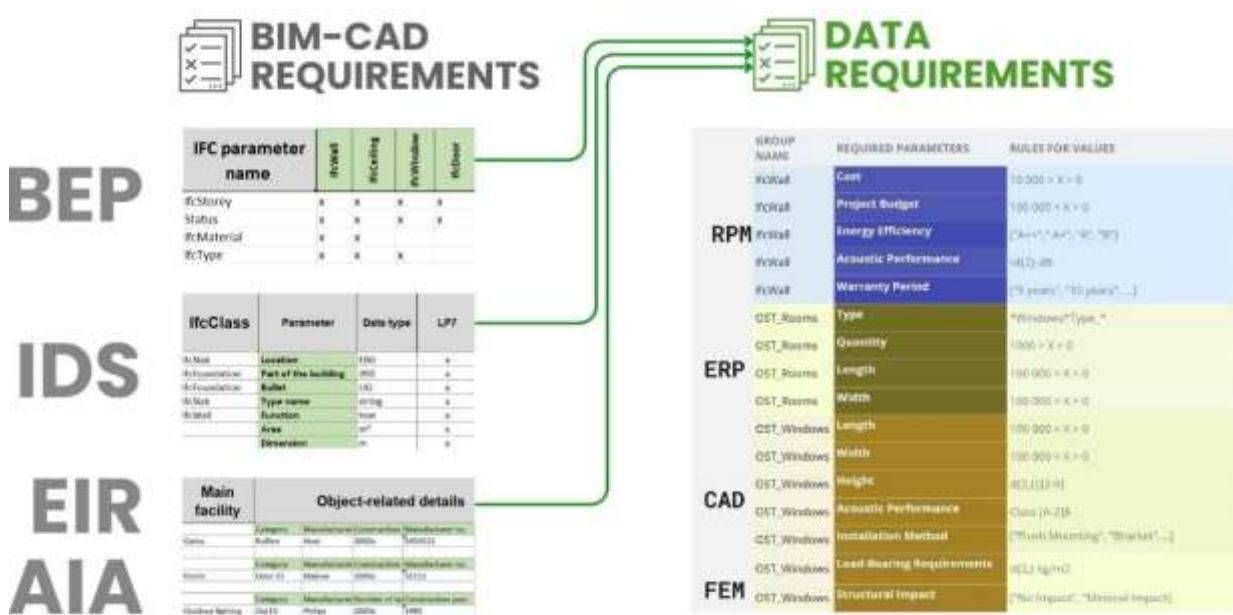
Zamiast upraszczać przetwarzanie danych, nowe koncepcje związane z danymi CAD (BIM) częściej generują dodatkowe zawiłości i spory już na etapie interpretacji i podstawowych definicji.

Jednym z najnowszych przykładów nowych koncepcji jest format IDS (wprowadzony w 2020 r.), który pozwala opisać wymagania dotyczące składu atrybutów modelu informacyjnego w otwartej koncepcji BIM. Wymagania IDS opisują informacje o atrybutach i ich wartościach granicznych w formie ustrukturyzowanej tabeli (Excel lub MySQL), która jest następnie tłumaczona na znaczniki pół-strukturalnego formatu XML, przemianowanego z XML na specjalny skrót IDS.

W przeciwieństwie do poglądu promowanego przez sprzedawców i wspieranego przez BIM i otwarty BIM, że obsługa danych w budownictwie jest wyjątkowa ze względu na wykorzystanie specjalistycznych narzędzi, takich jak CAD i BIM, formaty danych i praktyki zarządzania danymi w tej branży nie różnią się od tych w innych branżach

Liczبę wymagań dla projektów i formatów CAD (BIM) można uprościć, stosując pojedynczą tabelę wymagań z kolumnami atrybutów, opisaną szczegółowo w rozdziale "Tłumaczenie wymagań na formę ustrukturyzowaną", bez konieczności tłumaczenia pierwotnie ustrukturyzowanych wymagań na formaty inne niż tabelaryczne (IDS jest początkowo opisany za pomocą tabeli).

Uproszczone podejście (Rys. 6.3-2), które obejmuje kolumny dla identyfikatorów encji, właściwości i wartości brzegowych, które zostały szczegółowo omówione w poprzednich rozdziałach (Rys. 4.4-9, Rys. 4.4-16, Rys. 7.3-10), eliminuje potrzebę konwersji wymagań do formatu IDS-XML. Metoda ta zapewnia bezpośredni, mniej uciążliwy i bardziej przejrzysty mechanizm kontroli jakości danych. Opiera się na szeroko stosowanych narzędziach, od wyrażeń regularnych (RegEx) po ramki danych, Pandas i standardowe ETL -pipelines - dokładnie takie, jakie są używane przez profesjonalistów w innych sektorach gospodarki do pracy z danymi.



Rys. 6.3-2 Wymagania dotyczące danych w innych branżach są uproszczone do ustrukturyzowanego opisu atrybutów i ich wartości granicznych.

Z biegiem czasu w branży budowlanej, ze względu na zamknięty charakter danych, pojawia się coraz więcej nowych podejść i technik do kontrolowania i zarządzania tymi różnorodnymi danymi, chociaż dane w projektach budowlanych są zasadniczo takie same jak w innych dziedzinach. Podczas gdy inne branże z powodzeniem radzą sobie ze znormalizowanymi podejściami do przetwarzania danych, budownictwo nadal rozwija nowe i unikalne formaty danych, wymagania i koncepcje walidacji.

Metody i narzędzia wykorzystywane do gromadzenia, przygotowywania i analizowania danych w budownictwie nie powinny zasadniczo różnić się od tych stosowanych przez specjalistów w innych sektorach gospodarki.

Branża opracowała odrębny ekosystem terminologiczny, który wymaga krytycznej refleksji i ponownej oceny:

- Format STEP został umieszczony pod nową nazwą IFC, uzupełnioną o kategoryzację konstrukcji, bez uwzględnienia ograniczeń samego formatu STEP.
- Format parametryczny IFC jest wykorzystywany w procesach komunikacji danych pomimo braku jednolitego rdzenia geometrycznego wymaganego do wizualizacji i obliczeń.
- Dostęp do baz danych CAD -systemów jest promowany pod terminem "BIM ", bez dyskusji na temat specyfiki tych baz danych i dostępu do nich.
- Sprzedawcy promują interoperacyjność poprzez formaty IFC i USD, często nie stosując ich w praktyce, wykorzystując kosztowną inżynierię odwrotną, z którą sami się zmagają.
- Terminy LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD są powszechnie używane do opisania tych samych parametrów jednostki, bez odniesienia do narzędzi modelowania i weryfikacji od dawna stosowanych w innych branżach.

Branża budowlana pokazuje, że wszystkie powyższe, choć brzmi to dziwnie, są możliwe w branży budowlanej - zwłaszcza jeśli głównym celem jest monetyzacja każdego etapu przetwarzania danych poprzez sprzedaż specjalistycznych usług i oprogramowania. Z biznesowego punktu widzenia nie ma w tym nic złego. Kwestią otwartą pozostaje jednak to, czy takie akronimy i podejścia związane z CAD (BIM) rzeczywiście wnoszą wartość dodaną i upraszczają procesy zawodowe.

W branży budowlanej taki system działa, ponieważ sama branża osiąga większość swoich spekulacyjnych zysków w tym labiryncie systemów i akronimów. Firmy zainteresowane przejrzystymi procesami i otwartymi danymi są rzadkością. Ta złożona sytuacja będzie prawdopodobnie trwać w nieskończoność - dopóki klienci, inwestorzy, banki i fundusze private equity nie zaczną domagać się bardziej przejrzystego i świadomego podejścia do zarządzania informacjami.

Branża zgromadziła nadmierną liczbę akronimów, ale wszystkie one w różnym stopniu opisują te same procesy i wymagania dotyczące danych. Ich rzeczywista przydatność w upraszczaniu przepływu pracy pozostaje wątpliwa.

Podczas gdy koncepcje i akronimy marketingowe pojawiają się i znikają, same procesy walidacji wymagań dotyczących danych na zawsze pozostaną integralną częścią procesów biznesowych. Zamiast tworzyć coraz bardziej wyspecjalizowane formaty i regulacje, branża budowlana powinna szukać narzędzi, które już okazały się skuteczne w innych obszarach, takich jak finanse, przemysł i IT.

Mnogość terminów, akronimów i formatów tworzy iluzję dogłębnie opracowanych cyfrowych procesów konstrukcyjnych. Jednak koncepcje marketingowe i złożona terminologia często ukrywają prostą, ale niewygodną prawdę: dane pozostają trudno dostępne, słabo udokumentowane i sztywno powiązane z określonymi rozwiązaniami programowymi.

Aby wydostać się z tego błędного koła akronimów i formatów dla samych formatów, konieczne jest spojrzenie na systemy CAD (BIM) nie jak na magiczne narzędzia do zarządzania informacjami, ale jak na to, czym naprawdę są - wyspecjalizowanymi bazami danych. I to właśnie przez ten pryzmat można zrozumieć, gdzie kończy się marketing, a zaczyna prawdziwa praca z informacją.



ROZDZIAŁ 5.4.

PARAMETRYZACJA PROJEKTU I WYKORZYSTANIE LLM DO OBSŁUGI CAD

Iluzja unikalności danych CAD (BIM): droga do analityki i otwartych formatów

Nowoczesne platformy CAD (BIM) znacznie zmieniły podejście do zarządzania informacjami projektowymi i budowlanymi. Podczas gdy wcześniej narzędzia te były wykorzystywane głównie do tworzenia rysunków i modeli 3D, dziś funkcjonują one jako pełnoprawne repozytoria danych projektowych. Zgodnie z koncepcją pojedynczego źródła prawdy, model parametryczny coraz częściej staje się głównym, a często jedynym źródłem informacji o projekcie, zapewniając jego integralność i przydatność przez cały cykl życia projektu.

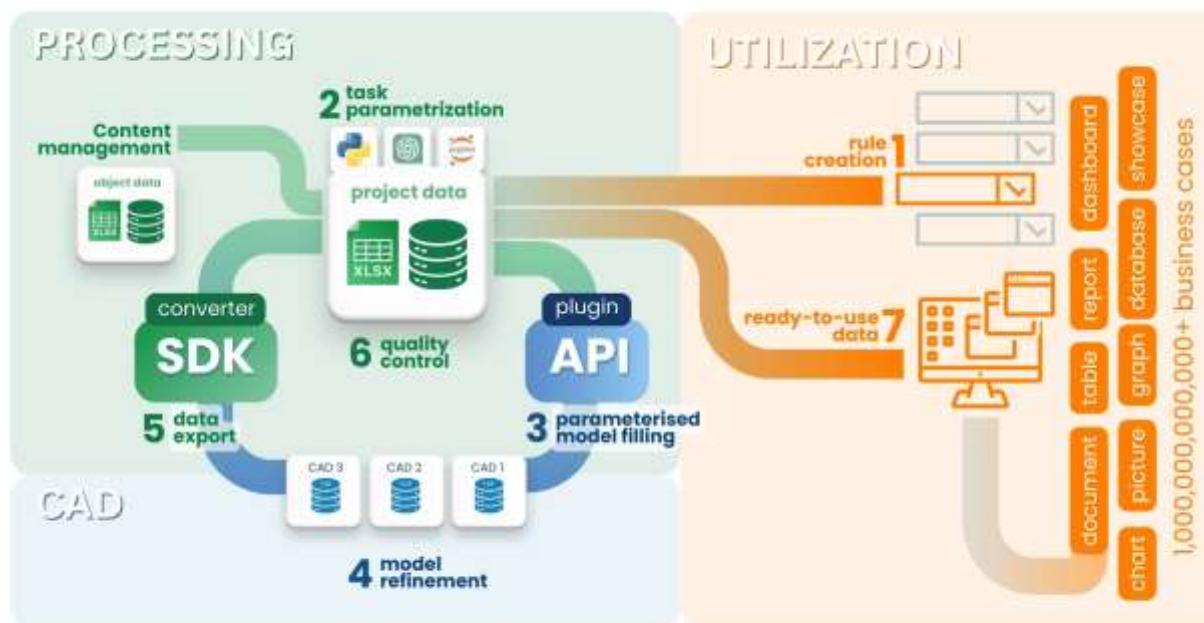
Kluczową różnicą między platformami CAD - (BIM-) a innymi systemami zarządzania danymi budowlanymi jest potrzeba specjalistycznych narzędzi i interfejsów API w celu uzyskania dostępu do informacji (jedynego źródła prawdy). Te bazy danych nie są uniwersalne w tradycyjnym sensie: zamiast otwartej struktury i elastycznej integracji, są zamkniętym środowiskiem, podłączonym na stałe do określonej platformy i formatu.

Pomimo złożoności pracy z danymi CAD istnieje ważniejsze pytanie, które wykracza poza techniczną realizację: czym tak naprawdę są bazy danych CAD (BIM)? Aby odpowiedzieć na to pytanie, konieczne jest wyjście poza zwykłe akronimy i pojęcia narzucone przez twórców oprogramowania. Zamiast tego warto skupić się na istocie pracy z informacją projektową: danych i ich przetwarzaniu.

Proces biznesowy w budownictwie rozpoczyna się nie od pracy w narzędziach CAD - lub BIM - ale od tworzenia wymagań projektowych i modelowania danych. Najpierw definiowane są parametry zadania: lista podmiotów, ich początkowe cechy i wartości graniczne, które należy wziąć pod uwagę przy rozwiązywaniu konkretnego zadania. Dopiero wtedy modele i elementy są tworzone w systemach CAD (BIM) na podstawie określonych parametrów

Proces poprzedzający tworzenie informacji w bazach danych CAD - (BIM-) jest całkowicie taki sam jak proces modelowania danych, który został szczegółowo omówiony w czwartej części książki i rozdziale "Modelowanie danych: model koncepcyjny, logiczny i fizyczny" (rys. 4.3-1).

Podobnie jak w modelowaniu danych tworzymy wymagania dla danych, które później chcemy przetwarzać w bazie danych, dla baz danych CAD menedżerowie tworzą wymagania projektowe w postaci kilku kolumn tabeli lub list par klucz-wartość (rys. 6.4-1, kroki 1-2). I tylko na podstawie tych początkowych parametrów przy użyciu API automatycznie lub ręcznie, projektant tworzy (lub raczej udoskonala) obiekty w bazach danych CAD- (BIM) (kroki 3-4), po czym są one ponownie sprawdzane pod kątem zgodności z początkowymi wymaganiami (kroki 5-6). Proces ten - definicja → tworzenie → walidacja → dostosowanie (kroki 2-6) - jest powtarzany iteracyjnie, aż jakość danych, podobnie jak w przypadku modelowania danych, osiągnie pożądany poziom dla systemu docelowego - dokumentów, tabel lub pulpitów nawigacyjnych (krok 7).

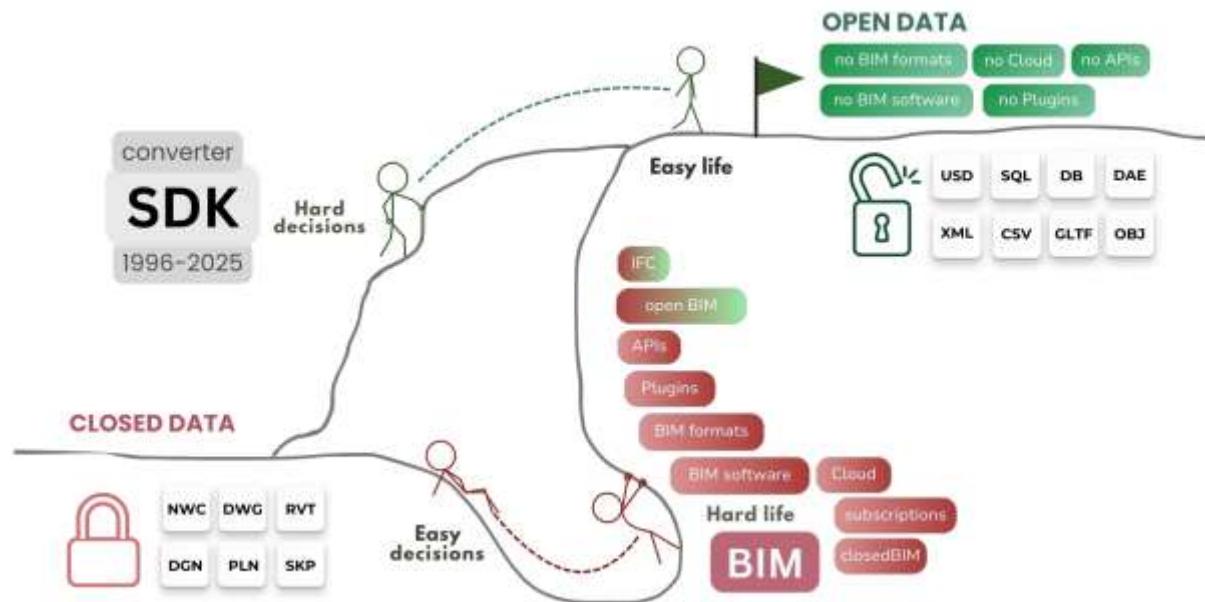


Rys. 6.4-1 Cykl nasycenia informacją baz danych dla procesów biznesowych w realizacji projektów budowlanych.

Jeśli rozważymy CAD (BIM) jako mechanizm przekazywania parametrów w postaci zestawu par klucz-wartość generowanych na podstawie wymagań zdefiniowanych poza środowiskiem projektowym (rys. 6.4-1, kroki 1-2), punkt ciężkości dyskusji przesuwa się z konkretnych rozwiązań programowych i ich ograniczeń na bardziej fundamentalne aspekty - strukturę danych, modele danych i wymagania dotyczące danych. Zasadniczo mówimy o nasyceniu parametrów bazy danych i klasycznym procesie modelowania danych (kroki 2-3 i 5-6). Jedyna różnica polega na tym, że ze względu na zamknięty charakter baz danych CAD i specyfikę stosowanych formatów, procesowi temu towarzyszy wykorzystanie specjalistycznych narzędzi BIM. Powstaje pytanie: na czym polega wyjątkowość BIM, skoro w innych branżach nie ma podobnych rozwiązań?

Przez ostatnie 20 lat BIM był pozycjonowany jako coś więcej niż tylko pojedyncze źródło danych. Pakiet CAD-BIM jest często sprzedawany jako narzędzie parametryczne z nieodłączną zintegrowaną bazą danych [64], zdolne do automatyzacji procesów projektowania, modelowania i zarządzania cyklem życia projektów budowlanych. Jednak w rzeczywistości BIM stał się bardziej narzędziem do utrzymywania użytkowników na platformie dostawców niż wygodną metodą zarządzania danymi i procesami.

W rezultacie dane CAD- (BIM-) są izolowane na ich platformach, ukrywając informacje o projekcie za zastrzeżonymi API i jądrami geometrii. Pozbawiło to użytkowników możliwości niezależnego dostępu do baz danych oraz wyodrębniania, analizowania, automatyzowania i przesyłania danych do innych systemów, z pominięciem ekosystemów dostawców.



Rys. 6.4-2 W budownictwie nowoczesne formaty wymagają zaawansowanych jąder geometrii, corocznie aktualizowanego API i specjalnych licencji na oprogramowanie CAD -(BIM -).

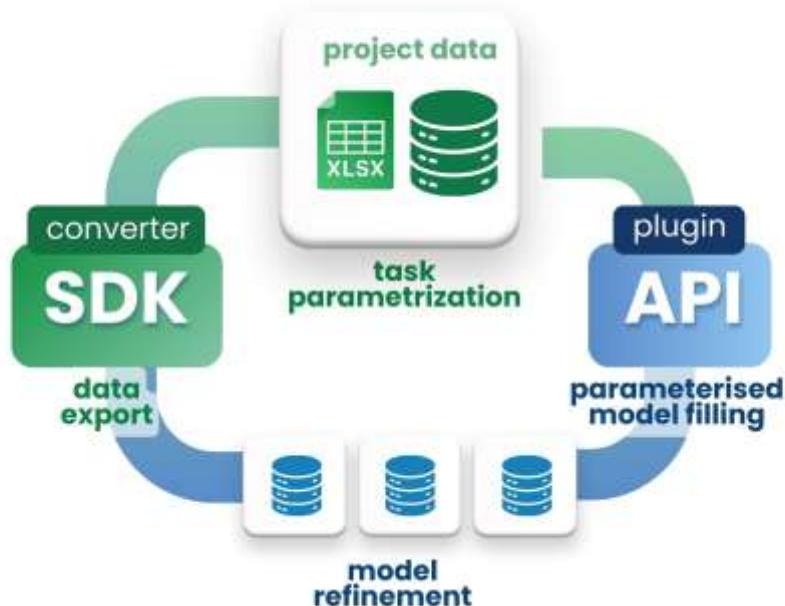
Firmy pracujące z nowoczesnymi narzędziami CAD powinny stosować to samo podejście do pracy z danymi, które w praktyce stosują wszyscy bez wyjątku producenci CAD: transformacja danych za pomocą SDK - narzędzia inżynierii odwrotnej, z którymi producenci CAD walczą od 1995 roku [75]. Mając pełny dostęp do bazy danych CAD i korzystając z narzędzi inżynierii odwrotnej, możemy uzyskać [127] płaski zbiór encji z atrybutami i wyeksportować je do dowolnego wygodnego otwartego formatu (rys. 6.4-2), obejmującego zarówno geometrię, jak i parametry elementów projektu. Takie podejście zasadniczo zmienia paradymat pracy z informacjami - od architektury zorientowanej na pliki do architektury zorientowanej na dane:

- Formaty danych takie jak RVT, IFC, PLN, DB1, CP2, CPIXML, USD, SQLite, XLSX, PARQUET i inne zawierają identyczne informacje o elementach tego samego projektu. Oznacza to, że znajomość konkretnego formatu i jego schematu nie powinna stanowić bariery w pracy z samymi danymi.
- Dane z dowolnych formatów mogą być łączone w jedną otwartą strukturę strukturalną i ziarnistą (Rys. 9.1-10) zawierającą geometrię trójkątną MESH i właściwości wszystkich jednostek obiektów, bez ograniczeń jąder geometrycznych.
- Analityka danych dąży do uniwersalności: korzystając z otwartych danych, możesz pracować z danymi projektu niezależnie od używanego formatu.
- Minimalizacja, a także zależność od interfejsów API i wtyczek dostawców: praca z danymi nie zależy już od umiejętności obsługi interfejsów API.

Gdy i CAD - wymagania dotyczące danych są przekształcane w łatwe do analizy ustrukturyzowane formaty reprezentacji - programiści nie są już zależni od określonych schematów danych i zamkniętych ekosystemów.

Projektowanie poprzez parametry: przyszłość CAD i BIM

Żaden projekt budowlany na świecie nigdy nie rozpoczął się w programie CAD. Zanim Rys. lub model nabierze kształtu w CAD, przechodzi przez etap koncepcyjny (rys. 6.4-1, etapy 1-2), gdzie nacisk kładziony jest na parametry, które definiują podstawową ideę i logikę przyszłego obiektu. Etap ten odpowiada poziomowi koncepcyjnemu w modelowaniu danych (rys. 4.3-6). Parametry mogą istnieć wyłącznie w umyśle projektanta, ale idealnie są one ułożone w formie ustrukturyzowanych list, tabel lub przechowywane w bazach danych (rys. 6.4-3), co pozwala na przejrzystość, powtarzalność i dalszą automatyzację procesu projektowania.



Rys. 6.4-3 Proces projektowania jest iteracyjnym procesem wypełniania bazy danych CAD informacjami z zewnątrz przy użyciu wymagań w łańcuchu wartości.

Przed rozpoczęciem samego modelowania CAD (logiczna i fizyczna faza modelowania danych (rys. 4.3-7)) ważne jest zdefiniowanie parametrów brzegowych, które służą jako podstawa projektu. Atrybuty te, podobnie jak inne wymagania, są zbierane z samego końca łańcucha wykorzystania danych (np. systemów) i dzięki nim ograniczenia, cele i kluczowe cechy przyszłych obiektów w projekcie są już zdefiniowane.

Samo modelowanie można w pełni zautomatyzować o 60-100% za pomocą narzędzi do modelowania parametrycznego (rys. 6.4-3), jeśli wymagania są dobrze zdefiniowane. Po opisaniu projektu w formie parametrów, jego tworzenie staje się technicznie wykonalne, np. za pomocą wizualnych języków programowania, takich jak Grasshopper Dynamo, wbudowanych w nowoczesne środowiska CAD lub darmowych rozwiązań w Blender, UE, Omniverse.



Rys. 6.4-4 Większość typowych projektów jest już dziś tworzona w pełni automatycznie dzięki narzędziom do programowania parametrycznego.

Już dziś duże projekty przemysłowe i typowe nie są tworzone rękami działu projektowego, ale za pomocą narzędzi parametrycznych i programowania wizualnego. Umożliwia to zbudowanie modelu opartego na danych, a nie na subiektywnych decyzjach konkretnego projektanta lub menedżera.

Treść poprzedza projekt. Projekt bez treści nie jest projektem, ale dekoracją [128].

- Jeffrey Zeldman, projektant stron internetowych i przedsiębiorca

Proces nie rozpoczyna się od rysowania lub modelowania 3D, ale od tworzenia wymagań. To właśnie wymagania określają, jakie elementy zostaną wykorzystane w projekcie, jakie dane muszą zostać przesłane do innych działów i systemów. Tylko istnienie ustrukturyzowanych wymagań umożliwia automatyczne sprawdzanie modeli na bieżąco (na przykład nawet co 10 minut bez odrywania projektanta od jego pracy).

Być może w przyszłości system CAD- (BIM-) stanie się tylko interfejsem do wypełniania bazy danych, a to, w którym narzędziu CAD modelowanie zostanie (poziom fizyczny), nie będzie miało znaczenia.

Podobnie w inżynierii mechanicznej, modelowanie 3D jest często stosowane, ale nie jest koniecznym lub obowiązkowym elementem projektu. W większości przypadków klasyczna dokumentacja 2D jest wystarczająca i na jej podstawie tworzony jest niezbędny model informacyjny. Model ten składa się z komponentów o strukturze zgodnej ze standardami branżowymi i zawiera wszystkie informacje niezbędne do zrozumienia projektu i organizacji produkcji. Model informacyjny fabryki jest następnie wykorzystywany do tworzenia modelu informacyjnego fabryki, do którego dodawane są określone produkty i schematy blokowe, które są już zorientowane na potrzeby technologów. Cały proces może być zorganizowany bez zbędnej złożoności, bez przeciążania systemu grafiką 3D tam, gdzie nie zapewnia ona rzeczywistych korzyści.

Ważne jest, aby zrozumieć, że sam model 3D i system CAD nie powinny odgrywać głównej roli - jest to tylko narzędzie do analizy ilościowej i geometrycznej. Wszystkie inne parametry, z wyjątkiem geometrii, które opisują jednostkę, powinny być przechowywane i przetwarzane poza środowiskiem CAD, jeśli to możliwe (BIM).

Projektowanie poprzez parametry to nie tylko trend, ale nieunikniona przyszłość branży budowlanej. Zamiat ręcznie tworzyć złożone modele 3D, projektanci będą pracować z danymi, weryfikować je i automatyzować procesy, zbliżając budownictwo do świata programowania. Z czasem procesy projektowania będą o parte na zasadach rozwoju oprogramowania:

- Utwórz wymagania → Utwórz model → Prześlij na serwer → Zatwierdź zmiany → Żądanie ściągnięcia
- Żądania ściągnięcia automatycznie uruchamiają kontrole modelu w odniesieniu do wymagań, które zostały utworzone przed lub w trakcie projektowania jako część żądania ściągnięcia.
- Po sprawdzeniu jakości danych i zatwierdzeniu, zmiany są wdrażane do projektu, wspólnej bazy danych lub automatycznie przesyłane do innych systemów.

Już teraz w inżynierii mechanicznej takie zmiany projektowe rozpoczynają się od utworzenia powiadomienia o zmianie. Podobny schemat czeka branżę budowlaną: projektowanie będzie procesem iteracyjnym, w którym każdy krok będzie wspierany przez wymagania parametryczne. Taki system pozwoli projektantom na tworzenie zautomatyzowanych kontroli i zautomatyzowanych żądań zmian dla określonych wymagań.

Projektant przyszłości jest przede wszystkim operatorem danych, a nie ręcznym modelarzem. Jego zadaniem jest wypełnienie projektu parametrycznymi obiekty, w których geometria jest tylko jednym z atrybutów.

To właśnie zrozumienie znaczenia modelowania danych, klasyfikacji i standaryzacji, które zostały szczegółowo omówione w poprzednich rozdziałach książki, odegra ważną rolę w transformacji. Przepisy projektowe przyszłości zostaną sformalizowane jako pary parametrów klucz-wartość w postaci schematów XLSX lub XML.

Przyszłość branży budowlanej to gromadzenie danych, ich analiza, walidacja i automatyzacja procesów przy użyciu narzędzi analitycznych. BIM (lub CAD) nie jest celem końcowym, a jedynie etapem ewolucji. Kiedy profesjonalisci zdadzą sobie sprawę, że mogą pracować bezpośrednio z danymi, omijając tradycyjne narzędzia CAD, sam termin "BIM" stopniowo ustąpi miejsca koncepcjom wykorzystania ustrukturyzowanych i granularnych danych projektu budowlanego.

Jednym z kluczowych czynników przyspieszających transformację było pojawienie się dużych modeli językowych (LLM) i opartych na nich narzędzi. Technologie te zmieniają sposób obsługi danych projektowych, umożliwiając dostęp do informacji bez konieczności dogłębnej znajomości interfejsów API lub rozwiązań dostawców. Dzięki LLM proces tworzenia wymagań i interakcji z danymi CAD staje się intuicyjny i przystępny.

Pojawienie się LLM w projektowaniu CAD procesy przetwarzania danych

Oprócz rozwoju narzędzi dostępu do baz danych CAD oraz otwartych i uproszczonych formatów CAD, pojawienie się narzędzi LLM (Large Language Models) rewolucjonizuje przetwarzanie danych projektowych. Podczas gdy w przeszłości dostęp do informacji odbywał się głównie za pośrednictwem złożonych interfejsów i wymagał umiejętności programowania oraz znajomości API, obecnie możliwa jest interakcja z danymi przy użyciu języka naturalnego.

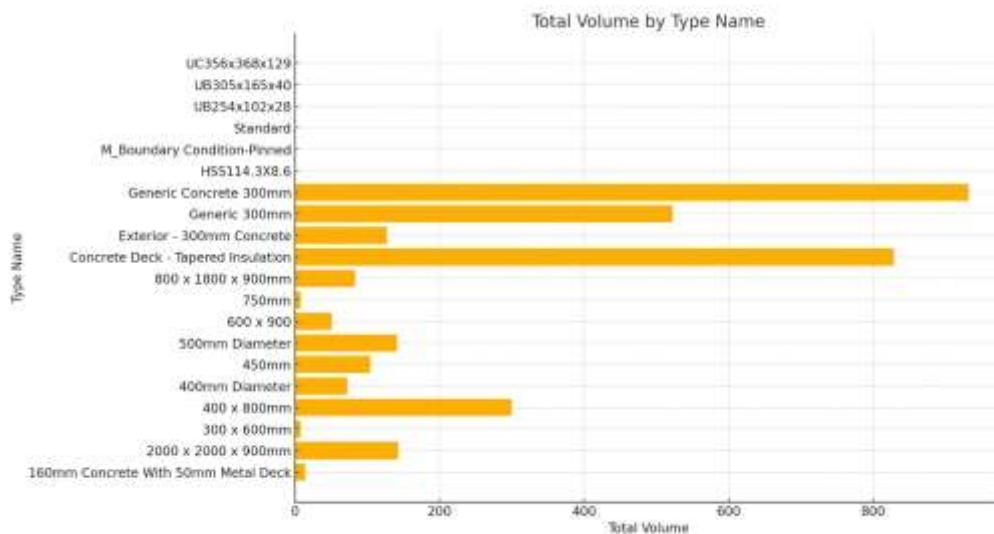
Inżynierowie, menedżerowie i planiści bez wykształcenia technicznego mogą uzyskać niezbędne informacje z danych projektowych, formułując zapytania w zwykłym języku. Pod warunkiem, że dane są ustrukturyzowane i dostępne (rys. 4.1-13), wystarczy zadać pytanie w czacie LLM typu: "Pokaż w tabeli z grupowaniem według typu wszystkie ściany o objętości większej niż 10 metrów sześciennych" - a model automatycznie przekształci to zapytanie w SQL lub kod w Pandas, generując tabelę podsumowującą, wykres lub nawet gotowy dokument.

Poniżej znajduje się kilka rzeczywistych przykładów interakcji modeli LLM z danymi projektowymi reprezentowanymi w różnych formatach CAD- (BIM-).

- ❷ Przykład zapytania w czacie LLM do projektu CAD w formacie RVT po konwersji (rys. 4.1-13) do tabelarycznej ramki danych (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN lub dowolnej innej):

Pogrupuj dane w ramce danych uzyskanej z pliku RVT według "nazwy typu" podczas sumowania parametru "objętość" i pokaż liczbę elementów w grupie. I pokaż to wszystko jako poziomy histogram bez wartości zerowych.

- Odpowiedź LLM jako poziomy wykres słupkowy (format PNG):



Rys. 6.4-5 Zamiast 17 kliknięć myszą lub 40 linii kodu za pomocą wtyczek, w LLM natychmiast pobieramy tabelę QTO za pomocą zapytania tekstowego.

- Aby wygenerować tabelę QTO typów ścian o łącznej powierzchni i ilości z kategorii "Ściany", sformułujmy zapytanie tekstowe dla LLM -chat:

Weź tylko te elementy z ramki danych projektu, które mają "OST_Walls" w parametrze "Category", pogrupuj je według "Type Name", zsumuj wartość kolumny "Area", dodaj ilość i wyświetl je w tabeli, usuwając wartości null.

- 💡 Odpowiedź LLM w formie gotowej tabeli QTO:

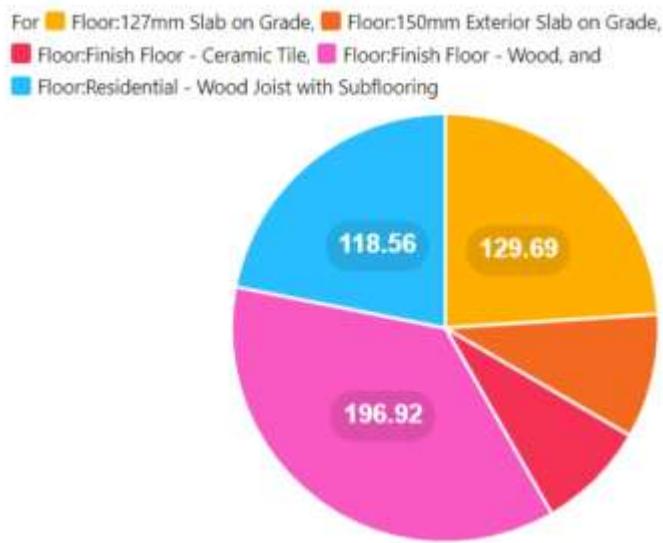
Type Name	Total Area	Count
CL_W1	393.12 sq m	10
Cavity wall_sliders	9.37 sq m	1
Foundation - 300mm Concrete	30.90 sq m	1
Interior - 165 Partition (1-hr)	17.25 sq m	3
Interior - Partition	186.54 sq m	14
Retaining - 300mm Concrete	195.79 sq m	10
SH_Curtain wall	159.42 sq m	9
SIP 202mm Wall - conc clad	114.76 sq m	4
Wall - Timber Clad	162.91 sq m	8

Rys. 6.4-6 Tworzenie tabeli QTO w języku naturalnym zapewnia taką samą jakość, jak w przypadku korzystania z narzędzi CAD - (BIM-).

- 💡 Zapytajmy o projekt w formacie IFC po przekonwertowaniu na tabelaryczną ramkę danych i wprowadźmy podobne zapytanie tekstowe do dowolnego czatu LLM:

Weź tylko te elementy z projektu, które mają wartości Level 1 i Level 2 w parametrze "Parent" i weź elementy, które mają wartości IfcSlab w parametrze "Category", a następnie pogrupuj te elementy według parametru "ObjectType", zsumuj wartości w parametrze "PSet_RVT _Dimensions Area" i pokaż je jako wykres kołowy.

■ Odpowiedź LLM jako gotowy wykres kołowy grup elementów z danych IFC :



Rys. 6.4-7 Wynikiem zapytania o dane IFC w formacie strukturalnym może być dowolny typ wykresu, który jest wygodny do zrozumienia danych.

Za każdym z gotowych rozwiązań (rys. 6.4-5 - rys. 6.4-7) kryje się kilkanaście linijek kodu Pythona wykorzystującego bibliotekę Pandas. Wynikowy kod można skopiować z czatu LLM i użyć w dowolnym lokalnym lub internetowym IDE, aby uzyskać identyczne wyniki poza czatem LLM.

Na tym samym czacie LLM możemy pracować nie tylko z projektami uzyskanymi z formatów 3D CAD (BIM), ale także z płaskimi rysunkami w formacie DWG, do których możemy wysłać zapytanie na czacie LLM, aby wyświetlić na przykład dane o grupach elementów w postaci linii lub geometrii 3D po konwersji do postaci ustrukturyzowanej.

Zautomatyzowana analiza plików DWG za pomocą LLM i Pandas

Proces przetwarzania danych z plików DWG - ze względu na nieustrukturyzowany charakter informacji - zawsze był złożonym zadaniem, wymagającym specjalistycznego oprogramowania i często ręcznej analizy. Jednak wraz z rozwojem sztucznej inteligencji i narzędzi LLM, stało się możliwe zautomatyzowanie wielu etapów tego, obecnie, głównie ręcznego procesu. Rozważmy prawdziwy potok żądań do LLM (w tym przykładzie ChatGPT) do pracy z rysunkami DWG, które umożliwiają pracę z projektem:

- Filtrowanie danych DWG według warstwy, ID i współrzędnych
- Wizualizacja geometrii elementów
- Automatyczne dodawanie adnotacji do rysunków na podstawie parametrów
- Rozwiń polilinie ściany do płaszczyzny poziomej
- Tworzenie interaktywnych wizualizacji 3D danych planarnych

■ Struktura i analiza danych konstrukcyjnych bez skomplikowanych narzędzi CAD

W naszym przypadku proces tworzenia Pipeline rozpoczyna się od sekwencyjnego generowania kodu za pomocą LLM. Najpierw generowane jest zapytanie opisujące zadanie. ChatGPT generuje kod Python, który jest wykonywany i analizowany, pokazując wynik wewnątrz pokoju rozmów. Jeśli wynik nie jest zgodny z oczekiwaniemi, zapytanie jest korygowane i proces jest powtarzany

Pipeline to sekwencja zautomatyzowanych kroków wykonywanych w celu przetwarzania i analizowania danych. W takim procesie każdy krok przyjmuje dane jako dane wejściowe, wykonuje transformacje i przekazuje wynik do następnego kroku.

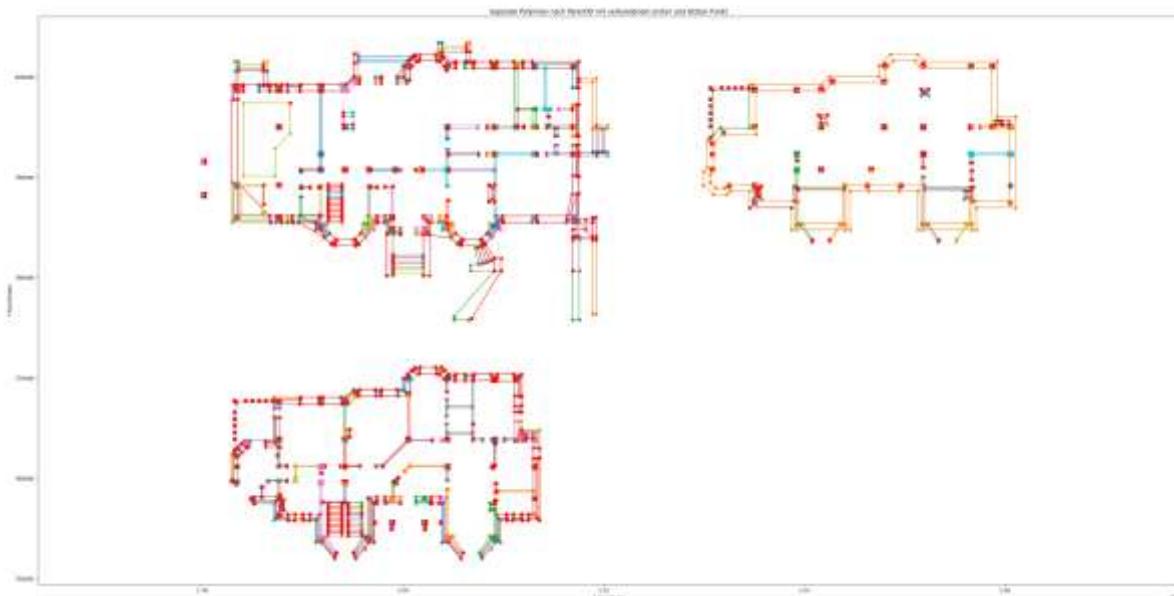
Po uzyskaniu pożądanego wyniku, kod jest kopowany z LLM i wklejany do kodu w postaci bloków w dowolnym z wygodnych IDE, w naszym przypadku na platformie Kaggle.com. Powstałe fragmenty kodu są łączone w jeden Pipeline, który automatyzuje cały proces - od ładowania danych do ich ostatecznej analizy. Takie podejście pozwala na szybki rozwój i skalowanie procesów analitycznych bez głębokiej wiedzy programistycznej. Pełny kod wszystkich poniższych fragmentów, wraz z przykładowymi zapytaniami, można znaleźć na platformie Kaggle.com, wyszukując "DWG Analyse with ChatGPT | DataDrivenConstruction". [129].

Proces pracy z danymi DWG, po konwersji do postaci strukturalnej (rys. 4.1-13), rozpoczęmy od klasycznego kroku - grupowania i filtrowania wszystkich danych rysunkowych, niezbędnych dla naszego zadania elementówściennych, a konkretnie polilinii (parametr "ParentID" pozwala na grupowanie linii w grupy), które w parametrze (kolumnie ramki danych) "Layer" mają wartość łańcuchową zawierającą następującą kombinację liter (RegEx) - "wall".

- ❷ Aby uzyskać kod dla podobnego zadania i wynik w postaci obrazka, należy napisać następujące zapytanie w LLM:

Najpierw sprawdź, czy ramka danych uzyskana z DWG zawiera zdefiniowane kolumny: 'Layer', 'ID', 'ParentID' i 'Point'. Następnie odfiltruj identyfikatory z kolumny "Layer", które zawierają ciąg "wall". Znajdź elementy w kolumnie "ParentID", które pasują do tych identyfikatorów. Zdefiniuj funkcję do czyszczenia i dzielenia danych w kolumnie "Point". Obejmuje to usunięcie nawiasów i podzielenie wartości na współrzędne "x", "y" i "z". Wykreśl dane za pomocą matplotlib. Dla każdego unikalnego "ParentID" narysuj oddzielną polilinię łączącą współrzędne "Point". Upewnij się, że pierwszy i ostatni punkt są połączone, jeśli to możliwe. Ustaw odpowiednie etykiety i tytuły, upewniając się, że osie x i y są jednakowo skalowane.

- ❸ Odpowiedź LLM da Ci gotowy obraz, za którym kryje się kod Pythona, który go wygenerował:



Rys. 6.4-8 LLM kod wyodrębnił wszystkie linie warstwy "ściana" z pliku DWG, wyczyścił ich współrzędne i skonstruował polilinie przy użyciu jednej z bibliotek Python.

- Teraz dodajmy do linii parametr obszaru, który każda polilinia ma w swoich właściwościach (w jednej z kolumn ramki danych):

Teraz pobierz tylko jeden "ParentID" z każdej polilinii - znajdź ten identyfikator w kolumnie "ID", weź wartość "Area", podziel przez 1 000 000 i dodaj tę wartość do wykresu.

- 2 Odpowiedź LLM wyświetli nowy wykres, na którym każda polilinia będzie miała podpis z jej obszarem:

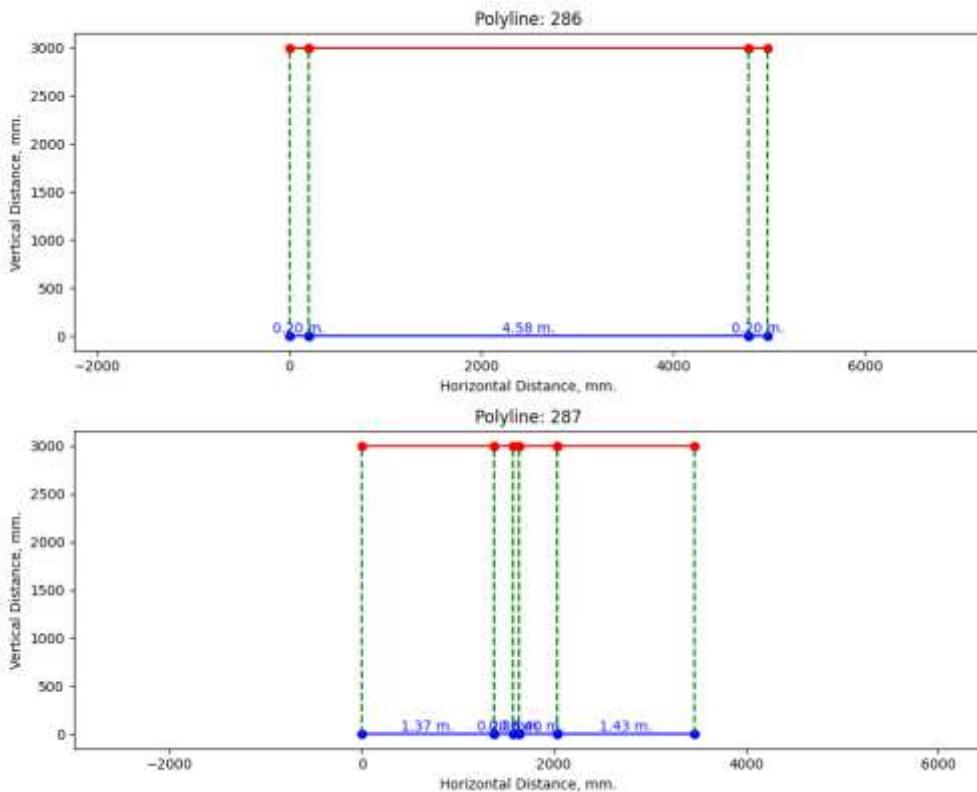


Rys. 6.4-9 LLM dodał kod, który pobiera wartości obszaru dla każdej polilinii i dodaje je do obrazu z wizualizacją linii.

- 3 Następnie przekształcimy każdą polilinię w linię poziomą, dodamy linię równoległą na wysokości 3000 mm i połączymy je w jedną płaszczyznę, aby w ten sposób pokazać układ powierzchni elementówściennych:

Należy pobrać wszystkie elementy z kolumny "Layer" z wartością "wall". Weź te identyfikatory jako listę z kolumny "ID" i znajdź te identyfikatory za całej ramki danych w kolumnie "ParentID". Wszystkie elementy są liniami, które są połączone w jedną polilinię. Każda linia ma inną geometrię x, y pierwszego punktu w kolumnie "Point". Musisz wziąć każdą polilinię po kolej i od punktu 0,0 poziomo narysować długość każdego segmentu z polilinii. długość każdego segmentu polilinii w jedną linię. Następnie narysuj dokładnie te same linie tylko 3000 wyżej, połącz wszystkie punkty w jedną płaszczyznę.

- 💡 Odpowiedź LLM wyświetli kod umożliwiający wykreślenie rysunków ścian na płaszczyźnie:

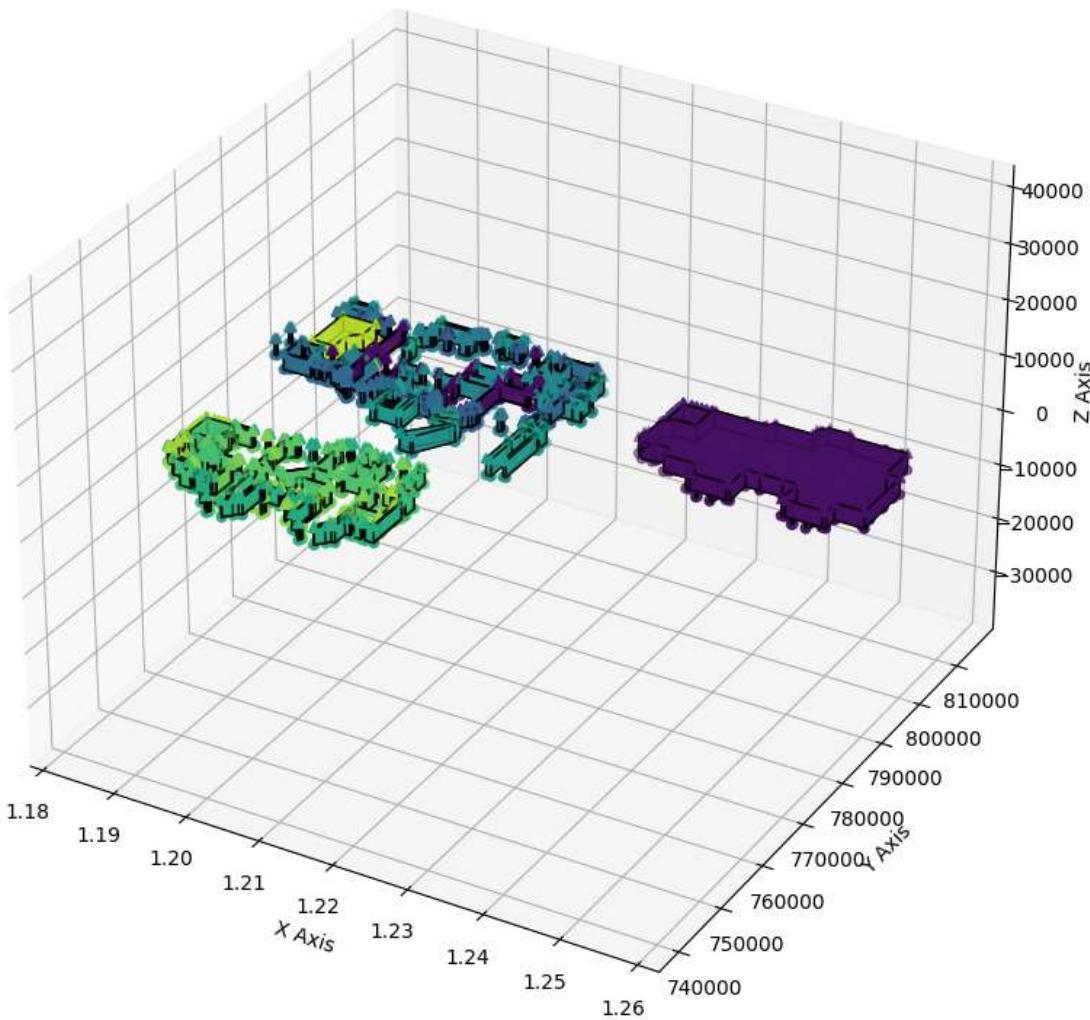


Rys. 6.4-10 Przekształcamy każdą polilinię za pomocą podpowiedzi w układ, który wizualizuje płaszczyzny ścian bezpośrednio na czacie LLM.

- 💡 Przejdźmy teraz od projekcji 2D do 3D - modelowanie ścian z płaskich linii poprzezłączenie górnej i dolnej warstwy polilinii:

Wizualizacja elementówściennych w 3D, łączenie polilinii na wysokościach $z = 0$ i $z = 3000$ mm. Aby utworzyć zamkniętą geometrię reprezentującą ściany budynku. Użyj narzędzi do tworzenia wykresów 3D Matplotlib.

- LLM wygeneruje interaktywny wykres 3D , w którym każda polilinia będzie reprezentowana jako zestaw płaszczyzn. Użytkownik będzie mógł swobodnie poruszać się między elementami za pomocą myszy komputerowej, eksplorując model w trybie 3D poprzez kopiowanie kodu z czatu do IDE:



Rys. 6.4-11 LLM pomógł stworzyć kod [129] do wizualizacji płaskich linii rysunkowych w widoku 3D , który można przeglądać w przeglądarce 3D w IDE.

Aby zbudować logiczny i powtarzalny Pipeline - od początkowej konwersji i załadowania pliku DWG do końcowego wyniku - zaleca się kopiowanie wygenerowanego bloku kodu LLM do IDE po każdym kroku. W ten sposób można nie tylko sprawdzić wynik na czacie, ale także natychmiast uruchomić go w środowisku programistycznym. Pozwala to na sekwencyjne budowanie procesu, debugowanie i dostosowywanie go w razie potrzeby.

Kompletny kod Pipeline wszystkich fragmentów (rysunki od 6.4-8 do 6.4-11) wraz z przykładowymi zapytaniami można znaleźć na platformie Kaggle.com, wyszukując "DWG Analyse with ChatGPT | DataDrivenConstruction". [129]. Na Kaggle można nie tylko przeglądać kod i używane podpowiedzi, ale także kopiować i testować cały Pipeline z oryginalnymi ramkami danych DWG w chmurze za darmo, bez

konieczności instalowania dodatkowego oprogramowania lub samego IDE.

Podejście przedstawione w tym rozdziale pozwala w pełni zautomatyzować sprawdzanie, przetwarzanie i generowanie dokumentów opartych na projektach DWG. Opracowany potok nadaje się zarówno do przetwarzania pojedynczych rysunków, jak i do przetwarzania wsadowego dziesiątek, setek i tysięcy plików DWG z automatycznym generowaniem niezbędnych raportów i wizualizacji dla każdego projektu.

Proces może być zorganizowany w sekwencyjny i przejrzysty sposób: najpierw dane z pliku CAD są automatycznie konwertowane do formatu XLSX, następnie ładowane do ramki danych, po czym następuje grupowanie, sprawdzanie i generowanie wyników - wszystko to jest zaimplementowane w pojedynczym notatniku Jupyter lub skrypcie Python, w dowolnym popularnym IDE. W razie potrzeby proces można łatwo rozszerzyć poprzez integrację z systemami zarządzania dokumentacją projektową: pliki CAD mogą być automatycznie pobierane zgodnie z określonymi kryteriami, wyniki mogą być zwracane z powrotem do systemu przechowywania, a użytkownicy mogą być powiadamiani, gdy wyniki są gotowe - za pośrednictwem poczty elektronicznej lub komunikatorów.

Korzystanie z czatów i agentów LLM do pracy z danymi projektowymi zmniejsza zależność od specjalistycznych programów CAD i umożliwia przeprowadzanie analiz i wizualizacji projektów architektonicznych bez konieczności ręcznej interakcji z interfejsem - bez klikania myszą i zapamiętywania złożonej nawigacji po menu.

Z każdym dniem branża budowlana będzie coraz częściej słyszeć o LLM, granularnych danych strukturalnych, DataFrames i kolumnowych bazach danych. Ujednolicone dwuwymiarowe DataFrames utworzone z różnych baz danych i formatów CAD, będą idealnym paliwem dla nowoczesnych narzędzi analitycznych, które są aktywnie wykorzystywane przez specjalistów z innych branż.

Sam proces automatyzacji zostanie znacznie uproszczony - zamiast studiować API zamkniętych niszowych produktów i pisać złożone skrypty do analizy lub przekształcania parametrów, teraz wystarczy sformułować zadanie w postaci zestawu pojedynczych poleceń tekstowych, które zostaną złożone w wymagany Pipeline lub Workflow-process dla wymaganego języka programowania, który działa bezpłatnie na prawie każdym urządzeniu. Koniec z oczekiwaniem na nowe produkty, formaty, wtyczki lub aktualizacje od dostawców narzędzi CAD- (BIM-). Inżynierowie i konstruktorzy będą mogli samodzielnie pracować z danymi przy użyciu prostych, bezpłatnych i łatwych do zrozumienia narzędzi, wspomaganych przez czaty i agentów LLM.

Kolejne kroki: przejście od zamkniętych formatów do otwartych danych

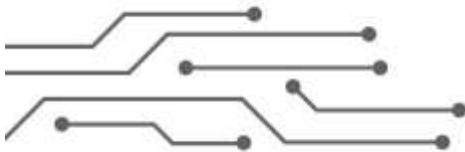
Podczas pracy z danymi projektowymi przyszłości jest mało prawdopodobne, aby ktokolwiek naprawdę musiał zrozumieć geometryczne jądra zastrzeżonych narzędzi lub nauczyć się setek niekompatybilnych formatów zawierających te same informacje. Jednak bez zrozumienia, dlaczego przejście na otwarte dane strukturalne jest ważne, trudno jest argumentować za wykorzystaniem nowych bezpłatnych narzędzi, otwartych danych i podejść, które prawdopodobnie nie będą promowane przez dostawców oprogramowania.

W tym rozdziale omówiliśmy kluczowe cechy danych CAD (BIM), ich ograniczenia i możliwości, a także fakt, że pomimo obietnic marketingowych producentów, inżynierowie i projektanci napotykają trudności w wyodrębnianiu, przesyłaniu i analizowaniu informacji projektowych każdego dnia. Zrozumienie architektury

tych systemów i poznanie alternatywnych podejść - opartych na otwartych formatach i automatyzacji poprzez LLM - może znacznie ułatwić życie nawet pojedynczemu profesjonalistie, nie mówiąc już o firmach. Podsumowując tę część, warto podkreślić główne praktyczne kroki, które pomogą zastosować podejścia omówione w tej sekcji w codziennych zadaniach:

- Rozszerz swój zestaw narzędzi do pracy z danymi projektu
 - Zapoznaj się z dostępnymi wtyczkami i narzędziami do wyodrębniania danych z używanych systemów CAD - (BIM-).
 - Zapoznaj się z dostępnymi zestawami SDK i interfejsami API, które pozwalają zautomatyzować wyodrębnianie danych z zamkniętych formatów bez konieczności ręcznego otwierania specjalistycznego oprogramowania.
 - Opanowanie podstawowych umiejętności pracy z otwartymi formatami geometrii nieparametrycznej (OBJ, glTF, USD, DAE) i odpowiadającymi im bibliotekami open source.
 - Postaraj się opracować system przechowywania metadanych projektu oddzielnie od geometrii poza rozwiązaniami CAD (BIM), aby uproszczyć analizę i integrację z innymi systemami.
 - Użyj LLM, aby zautomatyzować konwersję danych między formatami.
- Tworzenie własnych procesów obsługi informacji o projekcie
 - Zaczyna opisywać zadania i wymagania dotyczące modelowania za pomocą parametrów i ich wartości w prostych i ustrukturyzowanych formatach.
 - Utwórz osobistą bibliotekę skryptów lub bloków kodu dla często wykonywanych operacji.
- Promowanie korzystania z otwartych standardów w swojej pracy
 - Zaproś współpracowników i partnerów do udostępniania danych w otwartych formatach, które nie są ograniczone przez ekosystem dostawców oprogramowania.
 - Zademonstruj korzyści płynące z używania danych strukturalnych na konkretnych przykładach.
 - Inicjowanie dyskusji na temat problemów z zamkniętymi formatami i możliwych rozwiązań

Nawet jeśli nie możesz zmienić polityki swojej firmy w zakresie platform CAD - (BIM-), osobiste zrozumienie zasad pracy z danymi projektowymi w otwartych formatach pozwoli Ci znacznie zwiększyć wydajność Twojej pracy. Tworząc własne narzędzia i metody wyodrębniania i przekształcania danych z różnych formatów, nie tylko optymalizujesz przepływy pracy, ale także zyskujesz elastyczność, aby ominąć ograniczenia standardowych rozwiązań programowych.



VI CZĘŚĆ

PODEJMOWANIE DECYZJI W OPARCIU O DANE, ANALITYKA, AUTOMATYZACJA I UCZENIE MASZYNOWE

Siódma część skupia się na analizie danych i automatyzacji procesów w branży budowlanej. Omówiono w niej, w jaki sposób dane stają się podstawą do podejmowania decyzji i wyjaśniono zasady wizualizacji informacji w celu ich skutecznej analizy. Szczegółowo opisano kluczowe wskaźniki wydajności (KPI), metody oceny zwrotu z inwestycji (ROI) oraz tworzenie pulpitów nawigacyjnych do monitorowania projektów. Szczególną uwagę poświęcono procesom ETL (Extract, Transform, Load) i ich automatyzacji za pomocą potoków (Pipeline) w celu przekształcenia różnych danych w ustrukturyzowane informacje do analizy. Omówiono narzędzia do orkiestracji przepływu pracy, takie jak Apache Airflow, Apache NiFi i n8n, które umożliwiają tworzenie zautomatyzowanych potoków danych bez głębokiej wiedzy programistycznej. Dużą rolę odgrywają duże modele językowe (LLM) i ich wykorzystanie do uproszczenia analizy danych i automatyzacji rutynowych zadań

ROZDZIAŁ 6.1.

ANALITYKA DANYCH I PODEJMOWANIE DECYZJI W OPARCIU O DANE

Po etapach gromadzenia, strukturyzacji, czyszczenia i weryfikacji informacji powstał spójny i możliwy do analizy zbiór danych. Poprzednie części książki dotyczyły systematyzacji i strukturyzacji heterogenicznych źródeł - od dokumentów PDF i zapisów tekstowych spotkań po modele CAD i dane geometryczne. Szczegółowo opisano proces sprawdzania i dostosowywania informacji do wymagań różnych systemów i klasyfikatorów, eliminując duplikaty i niespójności.

Wszystkie obliczenia wykonywane na tych danych (trzecia, czwarta część książki) - od prostych przekształceń po obliczenia czasu, kosztów i wskaźników ESG (część piąta) - są zagregowanymi zadaniami analitycznymi. Stanowią one podstawę do zrozumienia aktualnego stanu projektu, oceny jego parametrów, a następnie podejmowania decyzji. W rezultacie dane, w wyniku obliczeń, zmieniają się ze zbioru rozproszonych rekordów w zarządzalny zasób zdolny do udzielenia odpowiedzi na kluczowe pytania biznesowe.

W poprzednich rozdziałach szczegółowo opisano procesy gromadzenia danych i kontroli jakości do wykorzystania w typowych przypadkach biznesowych i procesach specyficznych dla branży budowlanej. Analityka w tym kontekście jest pod wieloma względami podobna do zastosowań w innych branżach, ale ma wiele specyficznych cech.

W kolejnych rozdziałach szczegółowo opisany zostanie proces analizy danych, w tym etapy automatyzacji od początkowego pozyskania informacji i ich przekształcenia do późniejszego przeniesienia do docelowych systemów i dokumentów. Najpierw przedstawiona zostanie część teoretyczna, skupiająca się na wybranych aspektach podejmowania decyzji w oparciu o dane. Następnie, w kolejnych rozdziałach, przedstawiona zostanie część praktyczna związana z automatyzacją i budową ETL -Pipeline.

Dane jako zasób w procesie podejmowania decyzji

Podejmowanie decyzji w oparciu o dane jest często procesem iteracyjnym i rozpoczyna się od systematycznego gromadzenia informacji z różnych źródeł. Podobnie jak w naturalnym cyklu, poszczególne elementy danych i całe systemy informacyjne stopniowo wpadają do gleby - gromadząc się w repozytoriach informacji firm (Rys. 1.3-2). Z czasem dane te, podobnie jak opadłe liście i gałęzie, są przekształcane w cenny materiał. Grzybnia inżynierów i analityków danych organizuje i przygotowuje informacje do przyszłego wykorzystania oraz przekształca opadłe dane i systemy w cenny kompost, aby wyhodować nowe pędy i nowe systemy (rys. 1.2-5).

Trendy związane z powszechnym wykorzystaniem analityki w różnych branżach wyznaczają początek nowej ery, w której praca z danymi staje się podstawą działalności zawodowej (rys. 7.1-1). Ważne jest, aby specjaliści z branży budowlanej dostosowali się do tych zmian i byli gotowi wkroczyć w nową erę - erę danych i analityki.

Ręczne przenoszenie danych między tabelami i ręczne wykonywanie obliczeń stopniowo odchodzi w przeszłość, ustępując miejsca automatyzacji, analizie przepływu danych, analityce i uczeniu maszynowemu. Narzędzia te stają się kluczowymi elementami nowoczesnych systemów wspomagania decyzji.

W książce McKinsey "Rebooting. McKinsey's Guide to Overcoming Competition in the Age of Digital Technology and Artificial Intelligence" [130] przytoczono badanie przeprowadzone w 2022 r. z udziałem 1330 członków kadry kierowniczej wyższego szczebla z różnych regionów, branż i obszarów funkcjonalnych. [130] przytacza badanie przeprowadzone w 2022 r. z udziałem 1330 członków kadry kierowniczej wyższego szczebla z różnych regionów, branż i obszarów funkcjonalnych. Zgodnie z jego wynikami 70% liderów wykorzystuje zaawansowaną analitykę do generowania własnych pomysłów, a 50% wdraża sztuczną inteligencję w celu usprawnienia i automatyzacji procesów decyzyjnych.



Rys. 7.1-1 Analiza danych i analityka to główne narzędzie zwiększające szybkość podejmowania decyzji w firmie.

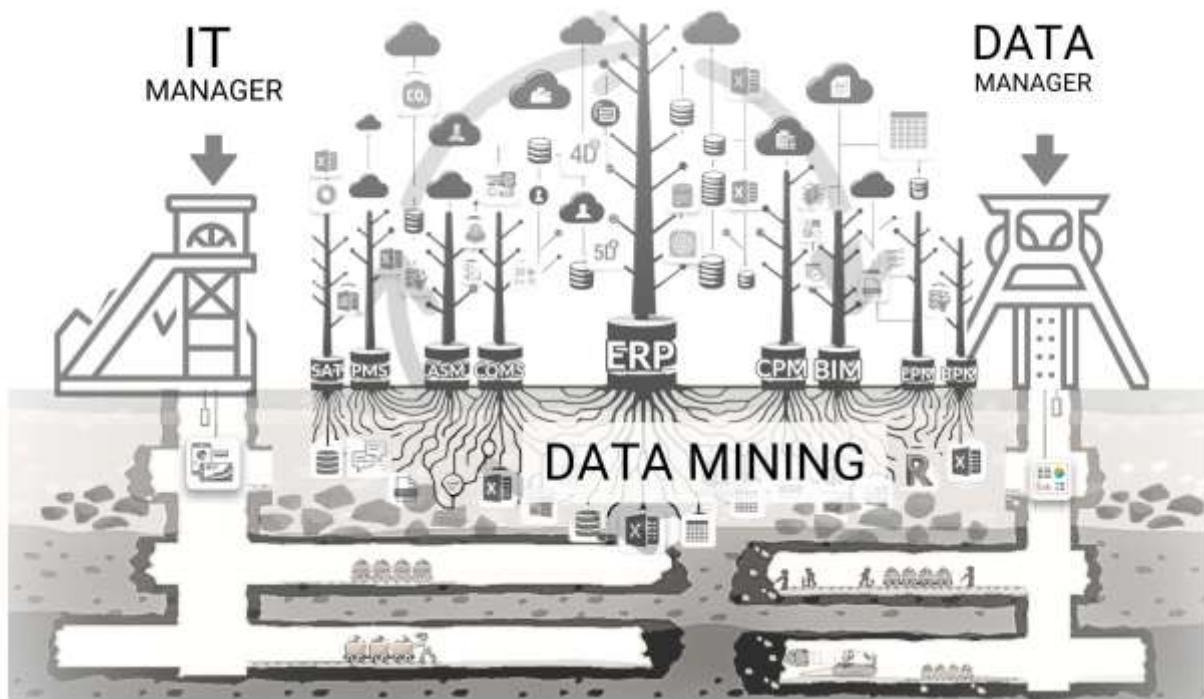
Analiza danych, podobnie jak rozprzestrzenianie się grzybni, penetruje próchnicę przeszłych decyzji, pomagając połączyć poszczególne systemy i prowadząc menedżerów do cennych spostrzeżeń. Wiedza ta, podobnie jak składniki odżywcze z rozkładających się drzew systemu danych, zasila nowe decyzje w firmie, prowadząc do skutecznych zmian i wzrostu jakości informacji, podobnie jak nowe pędy i kiełki wyłaniające się z bogatej i zdrowej gleby (Rys. 1.2-5).

Liczby mają ważną historię do opowiedzenia. Liczą na to, że dasz im jasny i przekonujący głos [131].

- Stephen Few, ekspert ds. wizualizacji danych

W średnich i małych firmach praca polegająca na wydobywaniu i przygotowywaniu informacji do dalszej analizy jest dziś niezwykle pracochłonnym procesem (Rys. 7.1-2), porównywalnym z XVIII-wiecznym wydobyciem węgla. Do niedawna praca nad eksploracją i przygotowaniem danych była raczej za-rezerwowana dla poszukiwaczy przygód pracujących w wysoce wyspecjalizowanej niszy z niewielkim i ograniczonym zestawem narzędzi do pracy z różnymi typami danych z nieustrukturyzowanych, luźno ustrukturyzowanych, mieszanych i zamkniętych źródeł.

Decydenci i menedżerowie często nie mają doświadczenia z heterogenicznymi danymi i systemami, a mimo to muszą podejmować decyzje na ich podstawie. W rezultacie podejmowanie decyzji w oparciu o dane w nowoczesnej branży budowlanej w ciągu ostatnich dziesięcioleci mniej przypominało zautomatyzowany proces, a bardziej wielodniową pracę ręczną górnika we wczesnych kopalniach węgla.



Rys. 7.1-2 W procesie eksploracji danych eksperci przechodzą przez złożoną ścieżkę przygotowania danych - od czyszczenia po strukturyzację do późniejszej analizy.

Podczas gdy nowoczesne metody pozyskiwania danych w branży budowlanej są z pewnością bardziej zaawansowane niż prymitywne techniki stosowane przez XII-wiecznych górników, nadal jest to zadanie złożone i obarczone wysokim ryzykiem, wymagające znacznych zasobów i wiedzy specjalistycznej, na które mogą sobie pozwolić tylko duże firmy. Procesy pozyskiwania i analizowania danych z nagromadzonego dziedzictwa poprzednich projektów były do niedawna podejmowane głównie przez duże, zaawansowane technologicznie firmy, które gromadziły i przechowywały dane konsekwentnie przez dziesięciolecia.

Wcześniej wiodącą rolę w analityce odgrywały dojrzałe technologicznie firmy, które gromadziły dane od dziesięcioleci. Dziś sytuacja się zmienia: dostęp do danych i narzędzi do ich przetwarzania staje się zdemokratyzowany - wcześniej skomplikowane rozwiązania są teraz dostępne dla każdego za darmo.

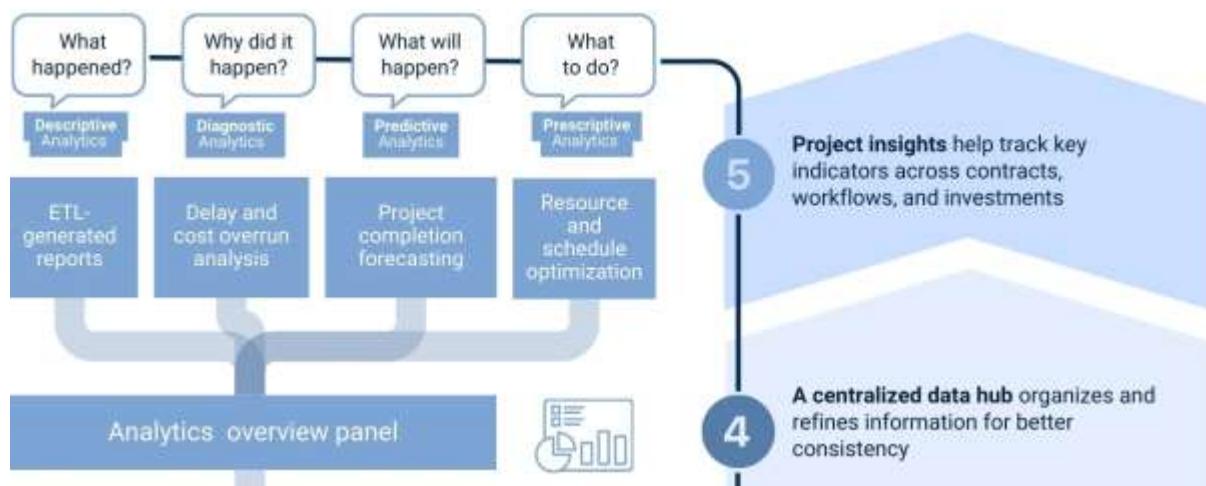
Zastosowanie analityki pozwala firmom podejmować dokładniejsze i bardziej świadome decyzje w czasie rzeczywistym. Poniższe studium przypadku ilustruje, w jaki sposób dane historyczne mogą pomóc w podejmowaniu rozsądnych decyzji finansowych:

- ⦿ **Kierownik projektu** - "Obecnie średnia cena betonu w mieście wynosi 82€ /m³, w kosztorysie mamy 95 €/m³".
- ⦿ **Estymator** - "W poprzednich projektach przekroczenie wynosiło około 15%, więc się wycofałem".
- ⦿ **Menedżer danych lub inżynier ds. kontroli po stronie klienta** - "Przyjrzymy się analizie ostatnich trzech przetargów".

Po przeanalizowaniu DataFrame z poprzednich projektów otrzymujemy:

- **Średnia rzeczywista cena zakupu:** 84,80 /m³
- **Średni wskaźnik przekroczenia wydatków:** +4,7%
- ⦿ **Zalecana stawka w kosztorysie:** ~ 85 /m³

Taka decyzja nie będzie już oparta na subiektywnych odczuciach, ale na konkretnych statystykach historycznych, co pomaga zmniejszyć ryzyko i zwiększyć ważność oferty przetargowej. Analiza danych z poprzednich projektów staje się swoistym "nawozem organicznym", z którego kiełkują nowe, trafniejsze rozwiązania.



Rys. 7.1-3 Analiza danych odpowiada na trzy kluczowe pytania: co się stało, dlaczego tak się stało i co należy zrobić dalej.

Decydenci i menedżerowie często stają przed koniecznością pracy z heterogenicznymi danymi i systemami bez wystarczającej wiedzy technicznej. W takich sytuacjach wizualizacja, jeden z pierwszych i najważniejszych kroków w procesie analitycznym, jest kluczową pomocą w zrozumieniu danych. Pozwala ona na przedstawienie informacji w wizualnej i zrozumiałej formie.

Wizualizacja danych: klucz do zrozumienia i podejmowania decyzji

W dzisiejszym przemyśle budowlanym, gdzie dane projektowe charakteryzują się złożonością i wielopoziomową strukturą, wizualizacja odgrywa kluczową rolę. Wizualizacja danych umożliwia kierownikom projektów i inżynierom wizualizację złożonych wzorców i trendów ukrytych w dużych, heterogenicznych ilościach danych.

Wizualizacja danych ułatwia zrozumienie statusu projektu: alokacji zasobów, trendów kosztowych lub zużycia materiałów. Wykresy i diagramy sprawiają, że złożone i suche informacje stają się dostępne i zrozumiałe, pozwalając szybko zidentyfikować kluczowe obszary, które wymagają uwagi i wykryć potencjalne problemy.

Wizualizacja danych nie tylko ułatwia interpretację informacji, ale jest kluczowym krokiem w procesie analitycznym i świadomym podejmowaniu decyzji zarządczych, pomagając odpowiedzieć na pytania "co się stało?" i "jak to się stało?" (Rys. 2.2-5). (Rys. 2.2-5).

Grafika jest wizualnym narzędziem do rozwiązywania problemów logicznych [132].

- Jacques Bertin, "Grafika i graficzne przetwarzanie informacji"

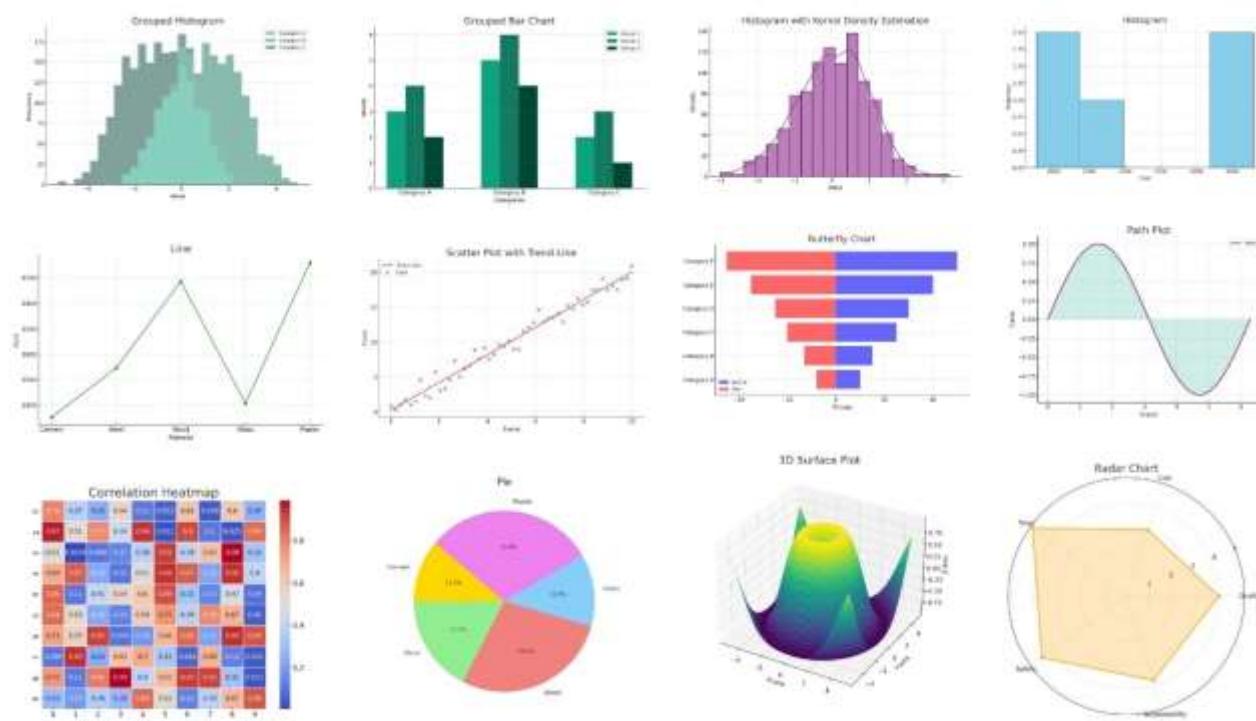
Przed podjęciem kluczowych decyzji kierownicy projektów częściej korzystają z wizualnych reprezentacji danych niż takich i trudnych do zinterpretowania liczb z arkuszy kalkulacyjnych lub wiadomości tekstowych.

Dane bez wizualizacji są jak materiały budowlane porozrzucane chaotycznie po placu budowy: ich potencjał jest niejasny. Dopiero gdy zostaną wyraźnie zwizualizowane, jak dom z cegieł i betonu, ich wartość staje się jasna. Dopóki dom nie zostanie zbudowany, nie można powiedzieć, czy stos materiałów stanie się małą chatką, luksusową willą czy wieżowcem.

Firmy posiadają dane z różnych systemów (Rys. 1.2-4 do rysunku 2.1-10), transakcje finansowe i obszerne dane tekstowe. Jednak wykorzystanie tych danych dla korzyści biznesowych jest często trudne. W takich sytuacjach wizualizacja staje się ważnym narzędziem do przekazywania znaczenia danych, pomagając przedstawić informacje w formatach zrozumiałych dla każdego eksperta, takich jak pulpity nawigacyjne, wykresy i diagramy.

Badanie PwC "What Students Need to Succeed in a Fast Changing Business World" (2015) podkreśla [9], że odnoszące sukcesy firmy wykraczają poza analizę danych i aktywnie wykorzystują interaktywne narzędzia wizualizacyjne, takie jak wykresy, infografiki i pulpity analityczne, aby wspierać podejmowanie decyzji. Według raportu - wizualizacja danych pomaga klientom zrozumieć historię opowiadaną przez dane za pomocą wykresów, diagramów, pulpitów nawigacyjnych i interaktywnych modeli danych.

Proces przekształcania informacji w wizualne formy graficzne, takie jak wykresy, grafy i diagramy, poprawia zrozumienie i interpretację danych przez ludzki mózg (Rys. 7.1-4). Pozwala to kierownikom projektów i analitykom na szybszą ocenę złożonych scenariuszy i podejmowanie świadomych decyzji w oparciu o rozpoznawalne wizualnie trendy i wzorce, a nie intuicję.



Rys. 7.1-4 Różne rodzaje wizualizacji mają na celu pomóc ludzkiemu mózgowi lepiej zrozumieć i nadać sens takim informacjom liczbowym.

Tworzenie wizualizacji z danych i korzystanie z różnych bezpłatnych bibliotek wizualizacji zostanie omówione bardziej szczegółowo w kolejnym rozdziale na temat procesów ETL.

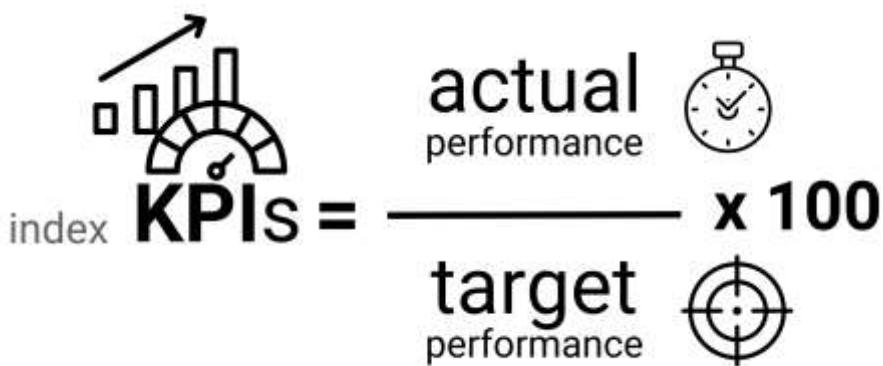
Wizualizacja staje się integralnym elementem pracy z danymi w branży budowlanej - pomaga nie tylko "zobaczyć" dane, ale także zrozumieć ich znaczenie w kontekście zadań zarządczych. Aby jednak wizualizacja była naprawdę przydatna, konieczne jest wcześniejsze określenie, co dokładnie należy wizualizować i które wskaźniki są naprawdę ważne dla oceny wydajności projektu. W tym miejscu do gry wkraczają wskaźniki wydajności, takie jak KPI i ROI. Bez nich nawet najpiękniejsze pulpity nawigacyjne mogą być tylko "szumem informacyjnym".

KPI i ROI

W dzisiejszej branży budowlanej zarządzanie wskaźnikami wydajności (KPI i ROI) oraz ich wizualizacja za pomocą raportów i pulpitów nawigacyjnych odgrywają kluczową rolę w poprawie wydajności i efektywności zarządzania projektami.

Jak w każdym biznesie, również w budownictwie konieczne jest jasne zdefiniowanie wskaźników, za pomocą których mierzy się sukces, zwrot z inwestycji i wydajność. Pozyskując dane na temat różnych procesów, organizacja oparta na danych musi najpierw nauczyć się identyfikować kluczowe **wskaźniki KPI (Key Performance Indicators)** - ilościowe mierniki, które odzwierciedlają stopień realizacji celów strategicznych i operacyjnych.

Aby obliczyć KPI, zwykle stosuje się wzór (Rys. 7.1-5), który obejmuje rzeczywiste i planowane wskaźniki. Na przykład, aby obliczyć indywidualny wskaźnik KPI dla projektu, pracownika lub procesu, należy podzielić rzeczywistą wydajność przez planowaną wydajność i pomnożyć wynik przez 100%.



Rys. 7.1-5 KPI są wykorzystywane do pomiaru sukcesu projektu lub procesu w osiąganiu kluczowych celów.

Na poziomie witryny można użyć bardziej szczegółowych wskaźników KPI:

- **Harmonogram kluczowych etapów** (fundamentowanie, instalacja, wykończenie) - pozwala kontrolować zgodność z planami prac.
- **Procent przekroczenia ilości materiałów** - pomaga zarządzać zaopatrzeniem i minimalizować straty.
- **Liczba nieplanowanych przestojów maszyn** - wpływa na produktywność i koszty.

Wybór niewłaściwych wskaźników może prowadzić do błędnych decyzji typu "co robić?" (Rys. 2.2-5). Na przykład, jeśli firma koncentruje się tylko na koszcie metra kwadratowego, ale nie bierze pod uwagę kosztów zmian, oszczędności na materiałach mogą prowadzić do gorszej jakości i wyższych kosztów w przyszłych projektach.

Podczas ustalania celów ważne jest, aby jasno określić, co jest mierzone. Niejasne sformułowania prowadzą do błędnych wniosków i komplikują kontrolę. Przyjrzyjmy się przykładom udanych i nieudanych wskaźników KPI w budownictwie.

Dobre wskaźniki KPI:

- ❑ "Do końca roku zmniejsz odsetek prac remontowych o 10 procent".
- ❑ "Zwiększenie szybkości montażu elewacji o 15 procent bez pogorszenia jakości do następnego kwartału"
- ❑ "Ograniczenie przestojów maszyn o 20% poprzez optymalizację harmonogramów pracy do końca roku"

Wskaźniki te są wyraźnie mierzalne, mają określone wartości i ramy czasowe.

Złe wskaźniki KPI:

- ❑ "Będziemy budować szybciej" (O ile szybciej? Co oznacza "szybciej"?).

- ❑ "Poprawimy jakość prac betoniarskich" (Jak dokładnie mierzona jest jakość?)
- ❑ "Poprawimy interakcję z wykonawcami na miejscu" (Jakie kryteria pokażą poprawę?)

Dobry KPI to taki, który można zmierzyć i obiektywnie ocenić. W budownictwie jest to szczególnie ważne, ponieważ bez jasnych wskaźników niemożliwe jest monitorowanie wydajności i osiąganie stabilnych wyników.

Oprócz KPI istnieje dodatkowa miara oceny efektywności inwestycji: **ROI (Return on Investment)** - wskaźnik zwrotu z inwestycji odzwierciedlający stosunek zysku do zainwestowanych środków. ROI pomaga ocenić, czy wprowadzenie nowych metod, technologii lub narzędzi jest uzasadnione: od rozwiązań cyfrowych i automatyzacji (np. Rys. 7.3-2) po wykorzystanie nowych materiałów budowlanych. Wskaźnik ten pomaga podejmować świadome decyzje dotyczące dalszych inwestycji w oparciu o ich rzeczywisty wpływ na rentowność biznesu

W kontekście zarządzania projektami budowlanymi, ROI (zwrot z inwestycji) może być wykorzystywany jako jeden z kluczowych wskaźników wydajności (KPI), jeśli celem firmy jest zmierzenie zwrotu z inwestycji w projekt, technologię lub usprawnienie procesu. Na przykład, jeśli wdrażana jest nowa technika zarządzania budową, ROI może pokazać, w jakim stopniu poprawiła ona rentowność.

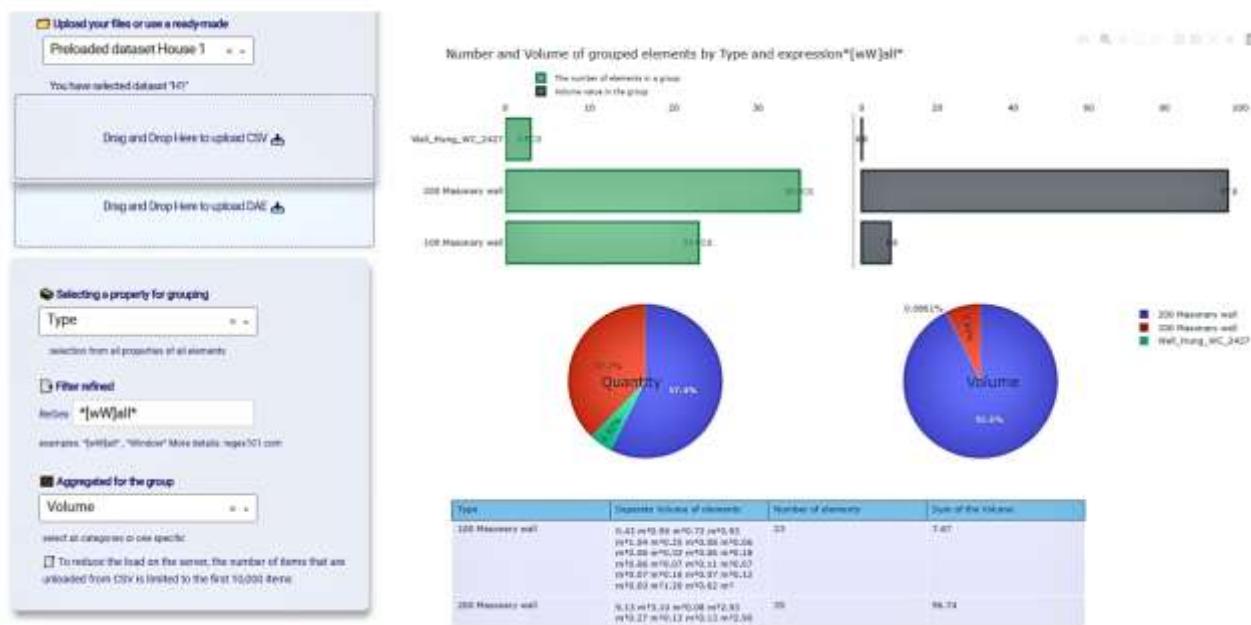
Regularne pomiary wskaźników KPI i ROI na podstawie danych zebranych z różnych źródeł, takich jak zużycie materiałów, godziny pracy i koszty, umożliwiają kierownictwu projektu efektywne zarządzanie zasobami i podejmowanie szybkich decyzji. Przechowywanie tych danych w dłuższej perspektywie pozwala na analizę przyszłych trendów i optymalizację procesów.

Różne wykresy i diagramy są wykorzystywane do wizualizacji wskaźników KPI, ROI i innych wskaźników, które są zwykle łączone w pulipy nawigacyjne.

Dashboardy i pulipy nawigacyjne: wizualizacja wskaźników w celu skutecznego zarządzania

Do wizualizacji wskaźników i metryk wykorzystywane są różne wykresy i diagramy, które są zwykle łączone w prezentacje danych i pulipy nawigacyjne. Pulipy te zapewniają skoncentrowany widok statusu projektu lub jego części, wyświetlając kluczowe wskaźniki (najlepiej w czasie rzeczywistym). Aktualne i stale aktualizowane pulipy pozwalają zespołowi szybko reagować na zmiany.

Dashboardy to narzędzia, które wizualizują oceny ilościowe, czyniąc je łatwo dostępnymi i zrozumiałymi dla wszystkich uczestników projektu.



Rys. 7.1-6 Zarządzanie wskaźnikami KPI i ich wizualizacja za pomocą pulpitów nawigacyjnych jest kluczem do poprawy produktywności i wydajności projektu.

Oto kilka przykładów popularnych narzędzi, w których można tworzyć pulpity nawigacyjne:

- **Power BI** to narzędzie firmy Microsoft do tworzenia interaktywnych raportów i pulpitów nawigacyjnych.
- **Tableau i Google Data Studio** to potężne narzędzia do wizualizacji danych i tworzenia pulpitów nawigacyjnych bez konieczności pisania kodu.
- **Plotly** (Rys. 7.1-6, Rys. 7.2-12) to biblioteka do tworzenia interaktywnych wykresów, a Dash to framework do tworzenia aplikacji internetowych do analizy danych. Mogą one być używane w połączeniu do tworzenia interaktywnych pulpitów nawigacyjnych.
- **Liczne biblioteki Pythona** (Rys. 7.2-9 - Rys. 7.2-11) - Python posiada wiele otwartych i darmowych bibliotek do wizualizacji danych, takich jak Matplotlib, Seaborn, Plotly, Bokeh i inne. Można ich używać do tworzenia wykresów i integrowania ich z aplikacją internetową przy użyciu frameworków takich jak Flask lub Django.
- **Biblioteki JavaScript**: umożliwia tworzenie interaktywnych pulpitów nawigacyjnych przy użyciu bibliotek JavaScript typu Open Source, takich jak D3.js lub Chart.js, i integrowanie ich z aplikacją internetową.

Do oceny wskaźników KPI i tworzenia pulpitów nawigacyjnych potrzebne są aktualne dane i jasny harmonogram gromadzenia i analizowania informacji.

Ogólnie rzecz biorąc, wskaźniki KPI, ROI i pulpity nawigacyjne w branży budowlanej stanowią podstawę analitycznego podejścia do zarządzania projektami. Pomagają one nie tylko monitorować i oceniać bieżący status, ale także dostarczają cennych informacji dla przyszłych procesów planowania i optymalizacji -

procesów, które zależą bezpośrednio od interpretacji danych i zadawania właściwych pytań w odpowiednim czasie.

Analiza danych i sztuka zadawania pytań

Interpretacja danych to ostatni etap analizy, w którym informacje nabierają sensu i zaczynają "przemawiać". To tutaj formułowane są odpowiedzi na kluczowe pytania: "co robić?" i "jak robić?" (Rys. 2.2-5). Ten etap pozwala podsumować wyniki, zidentyfikować wzorce, ustalić związki przyczynowo-skutkowe i wyciągnąć wnioski na podstawie wizualizacji i analizy statystycznej.

Być może już niedługo nadjdzie czas, gdy zdamy sobie sprawę, że aby w pełni stać się skutecznym obywatelem jednego z wielkich, nowych, złożonych państw świata, które obecnie się rozwijają, konieczna jest umiejętność liczenia, myślenia w kategoriach średnich, maksimów i minimów, tak jak obecnie konieczna jest umiejętność czytania i pisania [133].

- Samuel S. Wilkes, cytowany w przemówieniu prezydenckim z 1951 r. do Amerykańskiego Stowarzyszenia Statystycznego

Według raportu "Data Analytics and Artificial Intelligence in the Implementation of Government Projects" (2024) opublikowanego przez rząd Wielkiej Brytanii [83], wdrożenie analityki danych i sztucznej inteligencji (AI) może znacznie usprawnić procesy zarządzania projektami, zwiększając dokładność prognozowania czasu i kosztów, a także zmniejszając ryzyko i niepewność. W artykule podkreślono, że organizacje publiczne korzystające z zaawansowanych narzędzi analitycznych osiągają wyższą wydajność w inicjatywach infrastrukturalnych.

Nowoczesny biznes budowlany działający w wysoce konkurencyjnym i niskomarżowym środowisku czwartej rewolucji przemysłowej można porównać do operacji wojskowych. Tutaj przetrwanie i sukces firmy zależą od szybkości pozyskiwania zasobów i wysokiej jakości informacji - a tym samym od terminowego i świadomego podejmowania decyzji (rys. 7.1-7).

Jeśli wizualizacja danych jest "inteligencją", która zapewnia przegląd, to analityka danych jest "amunicją" potrzebną do działania. Odpowiada na pytania: *co robić?* i *jak to robić?*, tworząc podstawę do uzyskania przewagi konkurencyjnej na rynku.

Analityka przekształca różne dane w ustrukturyzowane i znaczące informacje, na których można opierać decyzje.

Zadaniem analityków i menedżerów jest nie tylko interpretowanie informacji, ale także oferowanie świadomych decyzji, identyfikowanie trendów, określanie relacji między różnymi typami danych i kategoryzowanie ich zgodnie z celami i specyfiką projektu. Korzystając z narzędzi do wizualizacji i metod analizy statystycznej, przekształcają oni dane w strategiczny zasób dla firmy.



Rys. 7.1-7 To analiza danych ostatecznie przekształca zebrane informacje w źródło do podejmowania decyzji.

Aby podejmować prawdziwie świadome decyzje w procesie analitycznym, konieczne jest nauczenie się, jak prawidłowo formułować pytania zadawane danym. Jakość tych pytań bezpośrednio wpływa na głębię uzyskanych spostrzeżeń, a w konsekwencji na jakość decyzji zarządczych.

Przeszłość istnieje tylko o tyle, o ile jest obecna w dzisiejszych zapisach. A to, co te zapisy reprezentują, zależy od pytań, które zadajemy. Nie ma innej historii niż ta [134].

- John Archibald Wheeler, fizyk, 1982

Sztuka zadawania głębokich pytań i krytycznego myślenia jest niezbędną umiejętnością w pracy z danymi. Większość ludzi ma tendencję do zadawania prostych, powierzchownych pytań, na które odpowiedź wymaga niewielkiego wysiłku. Jednak prawdziwa analiza zaczyna się od znaczących i przemyślanych pytań, które mogą odkryć ukryte relacje i związki przyczynowo-skutkowe w informacjach, które mogą być ukryte za wieloma warstwami rozumowania.

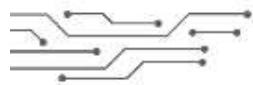
Według badania "Data-Driven Transformation: Accelerating at Scale Now" (BCG, 2017) [135], u-dana transformacja cyfrowa wymaga inwestycji w możliwości analityczne, programy zarządzania zmianą oraz dostosowania celów biznesowych do inicjatyw IT. Firmy, które tworzą kulturę opartą na danych, powinny inwestować w możliwości analizy danych i uruchamiać programy zarządzania zmianą, aby zaszczepić nowe myślenie, zachowania i sposoby pracy.

Bez inwestycji w rozwój kultury analitycznej, ulepszania narzędzi danych i szkolenia specjalistów, firmy będą nadal ryzykować podejmowanie decyzji w oparciu o nieaktualne lub niekompletne informacje - lub polegając na subiektywnych opiniach menedżerów HiPPO (Rys. 2.1-9).

Uświadomienie sobie istotności i potrzeby ciągłego aktualizowania analiz i pulpitów nawigacyjnych nieuchronnie prowadzi kierownictwo do zrozumienia znaczenia automatyzacji procesów analitycznych. Automatyzacja zwiększa szybkość podejmowania decyzji, zmniejsza wpływ czynnika ludzkiego i zapewnia trafność danych. Wraz z wykładniczym wzrostem ilości informacji, szybkość staje się nie tylko przewagą konkurencyjną, ale kluczowym czynnikiem trwałego sukcesu.

Automatyzacja procesów analizy i przetwarzania danych jest nierozerwalnie związana z tematem ETL (Extract, Transform, Load). Podobnie jak w procesie automatyzacji musimy przekształcić dane, w procesie

ETL dane są pobierane z różnych źródeł, przekształcane zgodnie z niezbędnymi wymaganiami i ładowane do systemów docelowych w celu dalszego wykorzystania.



ROZDZIAŁ 6.2.

PRZEPŁYW DANYCH BEZ RĘCZNEGO WYSIŁKU: DLACZEGO ETL JEST POTRZEBNY

Automatyzacja ETL: redukcja kosztów i przyspieszenie obsługi danych

Kiedy kluczowe wskaźniki wydajności (KPI) przestają rosnąć pomimo wzrostu ilości danych i wielkości zespołu, kierownictwo firmy nieuchronnie zdaje sobie sprawę z potrzeby automatyzacji procesów. Prędzej czy później świadomość ta staje się bodźcem do uruchomienia kompleksowej automatyzacji, której głównym celem jest zmniejszenie złożoności procesów, przyspieszenie przetwarzania i zmniejszenie zależności od czynnika ludzkiego.

Według badania McKinsey "How to Build a Data Architecture to Drive Innovation - Today and Tomorrow" (2022) [136], firmy korzystające z architektury danych strumieniowych zyskują znaczną przewagę, ponieważ mogą analizować informacje w czasie rzeczywistym. Technologie strumieniowe umożliwiają bezpośrednią analizę komunikatów w czasie rzeczywistym i zastosowanie konserwacji predykcyjnej w produkcji poprzez analizę danych z czujników w czasie rzeczywistym.

Uproszczenie procesów to automatyzacja, w której tradycyjne funkcje manualne są zastępowane przez algorytmy i systemy.

Kwestia automatyzacji, a raczej "minimalizacji roli człowieka w przetwarzaniu danych", jest nieodwracalnym i bardzo wrażliwym procesem dla każdej firmy. Specjalisci w dowolnej dziedzinie zawodowej często wahają się przed pełnym ujawnieniem swoich metod i subtelności pracy innym optymalizatorom, zdając sobie sprawę z ryzyka utraty pracy w szybko ewoluującym środowisku technologicznym.

Jeśli chcesz narobić sobie wrogów, spróbuj coś zmienić [137].

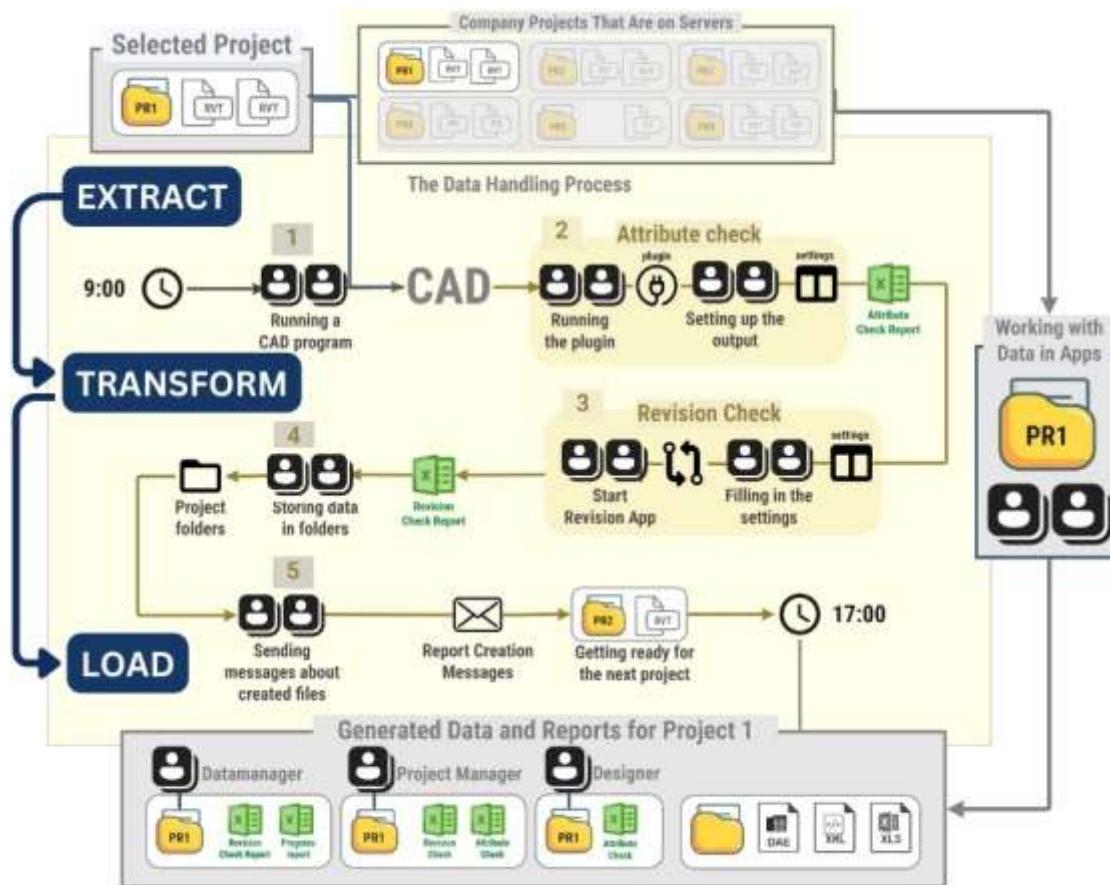
- Woodrow Wilson, przemówienie do kongresu sprzedawców, Detroit, 1916 r.

Pomimo oczywistych korzyści płynących z automatyzacji, wiele firm nadal ma wysoki udział pracy ręcznej w swoich codziennych praktykach, zwłaszcza w obszarze danych inżynieryjnych. Aby zilustrować obecną sytuację, przyjrzyjmy się typowemu przykładowi sekwencyjnego przetwarzania danych w ramach takich procesów.

Ręczną obsługę danych można zilustrować na przykładzie interakcji z informacjami uzyskanymi z baz danych CAD. Tradycyjne przetwarzanie danych ("ręczny" ETL -process) w działach CAD (BIM) w celu tworzenia tabel atrybutów lub tworzenia dokumentacji na podstawie danych projektowych odbywa się w następującej kolejności (rys. 7.2-1):

1. **Ręczna ekstrakcja (Extract):** użytkownik ręcznie otwiera projekt - uruchamiając aplikację CAD (BIM) (rys. 7.2-1 krok 1).

2. **Weryfikacja:** kolejny krok zwykle obejmuje ręczne uruchomienie kilku wtyczek lub aplikacji pomocniczych w celu przygotowania danych i oceny ich jakości (rys. 7.2-1 krok 2-3).
3. Ręczna **transformacja (Transform):** po przygotowaniu rozpoczyna się przetwarzanie danych, które wymaga ręcznej obsługi różnych narzędzi programowych, w których dane są przygotowywane do przesłania (rys. 7.2-1 krok 4).
4. Ręczne **przesyłanie (Load):** ręczne przesyłanie przekonwertowanych danych do zewnętrznych systemów, formatów danych i dokumentów (rys. 7.2-1 krok 5).



Rys. 7.2-1 Tradycyjne ręczne przetwarzanie ETL jest ograniczone pragnieniami i fizycznymi możliwościami poszczególnych techników.

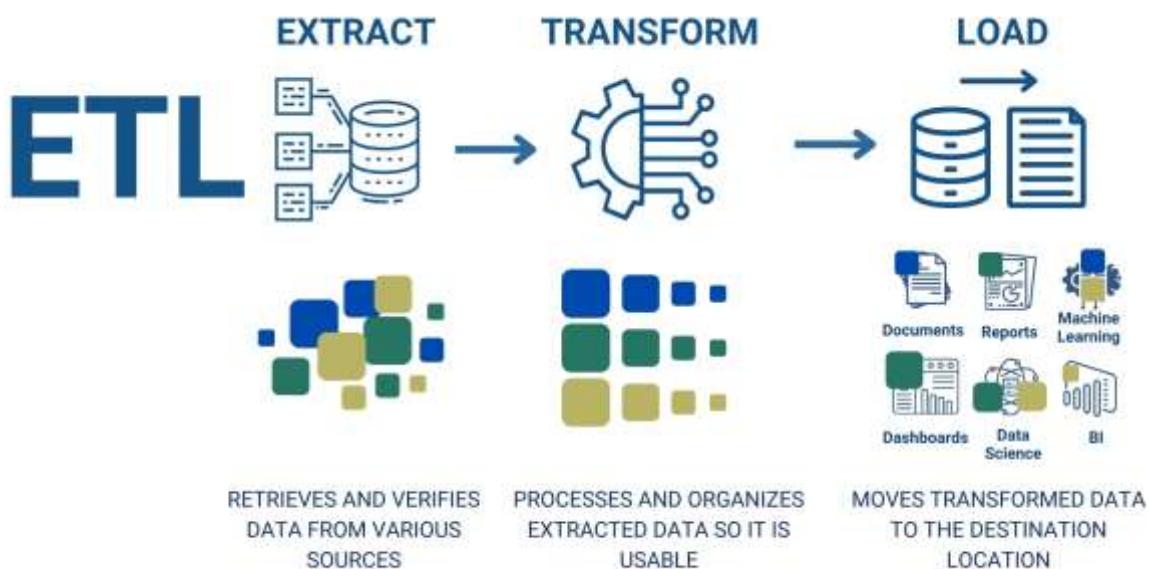
Taki przepływ pracy jest przykładem klasycznego ETL -procesu - ekstrakcji, transformacji i ładowania (ETL). W przeciwieństwie do innych branż, w których automatyczne potoki ETL od dawna są standardem, branża budowlana jest nadal zdominowana przez pracę ręczną, co spowalnia procesy i zwiększa koszty

ETL (Extract, Transform, Load) to proces pozyskiwania danych z różnych źródeł, przekształcania ich do pożądanego formatu i ładowania do systemu docelowego w celu dalszej analizy i wykorzystania.

ETL to proces, który obejmuje trzy kluczowe elementy przetwarzania danych: Wyodrębnianie, Przekształcanie i Ładowanie (Rys. 7.2-2):

- **Extract** - wyodrębnia dane z różnych źródeł (pliki, bazy danych, API).
- **Transform** - czyszczenie danych, agregacja, normalizacja i przetwarzanie logiczne.
- **Załaduj** - załaduj ustrukturyzowane informacje do hurtowni danych, raportu lub systemu BI.

Wcześniej w książce koncepcja ETL była poruszana tylko sporadycznie: w konwersji nieustrukturyzowanego zeskanowanego dokumentu do ustrukturyzowanego formatu tabelarycznego (rys. 4.1-1), w kontekście formalizacji wymagań w celu usystematyzowania postrzegania zarówno życia, jak i procesów biznesowych (rys. 4.4-20) oraz w automatyzacji walidacji danych i przetwarzania danych z rozwiązań CAD. Przyjrzymy się teraz ETL bardziej szczegółowo w kontekście typowych przepływów pracy.



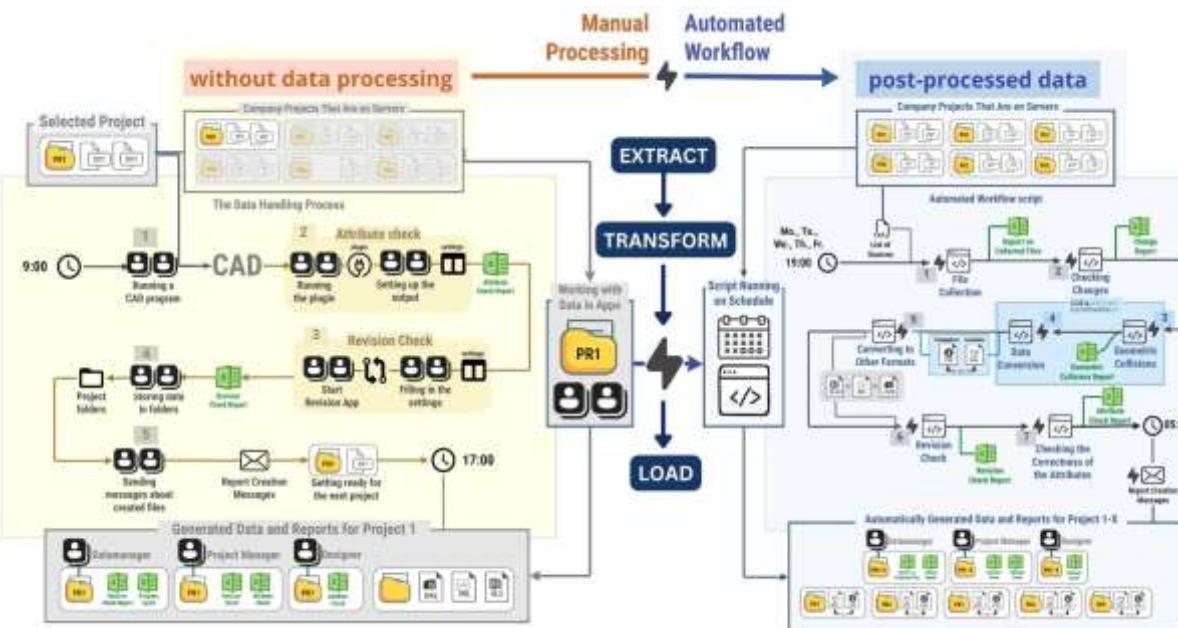
Rys. 7.2-2 ETL automatyzuje powtarzalne zadania przetwarzania danych.

Ręczny lub półautomatyczny proces ETL wymaga menedżera lub technika, który ręcznie zarządza wszystkimi etapami - od gromadzenia danych po generowanie raportów. Taki proces zajmuje dużo czasu, zwłaszcza gdy godziny pracy są ograniczone (np. od 9:00 do 17:00).

Firmy często starają się rozwiązać problem niskiej wydajności i niskiej prędkości poprzez zakup modułowych zintegrowanych rozwiązań (ERP, PMIS, CPM, CAFM, itp.), które są następnie dalej rozwijane przez zewnętrznych sprzedawców i konsultantów. Jednak ci sprzedawcy i zewnętrzni deweloperzy często stają się krytycznym punktem zależności: ich ograniczenia techniczne bezpośrednio wpływają na wydajność całego systemu i całej firmy, jak opisano szczegółowo w poprzednich rozdziałach dotyczących zastrzeżonych systemów i formatów. Problemy wynikające z fragmentacji i zależności zostały szczegółowo omówione w rozdziale "Jak firmy budowlane toną w chaosie danych".

Jeśli firma nie jest gotowa na wdrożenie dużej modułowej platformy od jednego z dostawców, zaczyna szukać alternatywnych sposobów automatyzacji. Jednym z nich jest opracowanie własnego modułowego otwartego systemu ETL -conveyors, w którym każdy etap (ekstrakcja, transformacja, walidacja, ładowanie) jest realizowany jako skrypty wykonywane zgodnie z harmonogramem.

W zautomatyzowanej wersji tego samego przepływu pracy ETL (rys. 7.2-1) przepływ pracy wygląda jak kod modułowy, który rozpoczyna się od przetwarzania danych i tłumaczenia ich na otwartą formę strukturalną. Po otrzymaniu ustrukturyzowanych danych różne skrypty lub moduły są uruchamiane automatycznie, zgodnie z harmonogramem, w celu sprawdzenia zmian, przekształcenia i wysłania wiadomości (rys. 7.2-3).



Rys. 7.2-3 Po lewej obróbka ręczna, po prawej proces automatyczny, który w przeciwieństwie do tradycyjnej obróbki ręcznej nie jest ograniczony możliwościami użytkownika.

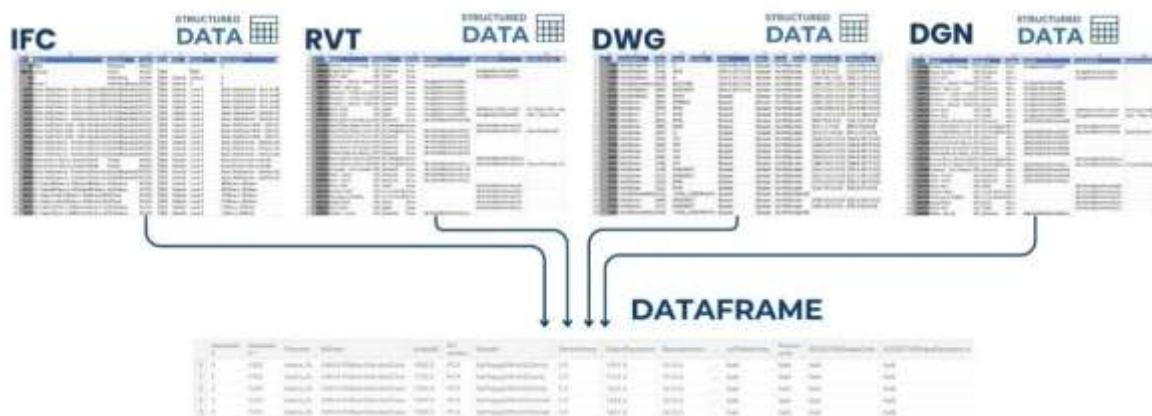
W zautomatyzowanym przepływie pracy przetwarzanie danych jest uproszczone dzięki wstępнемu przetwarzaniu danych ET(L): strukturyzacji i ujednoliceniu.

W tradycyjnych metodach przetwarzania specjalisci pracują z danymi "takimi, jakie są" - pobranymi z systemów lub oprogramowania. Z kolei w zautomatyzowanych procesach dane często najpierw przechodzą przez ETL -payplane, gdzie są doprowadzane do spójnej struktury i formatu odpowiedniego do dalszego wykorzystania i analizy.

Weźmy praktyczny przykład ETL, demonstrujący proces walidacji tabeli danych opisany w rozdziale "Walidacja danych i wyniki walidacji" (Rys. 4.4-13). Aby to zrobić, używamy biblioteki Pandas w połączeniu z LLM do zautomatyzowanej analizy danych i procesów przetwarzania.

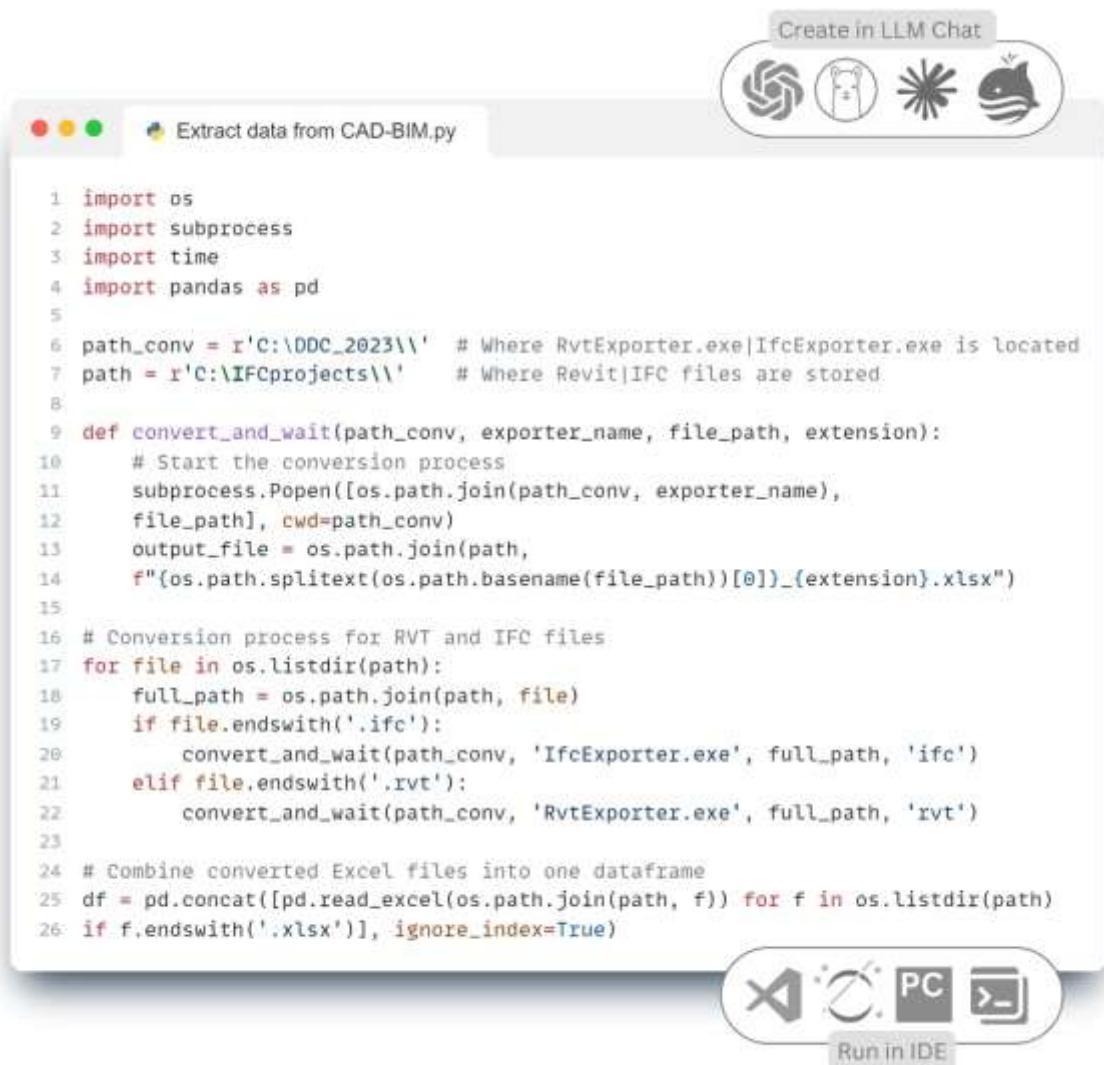
ETL Extract: gromadzenie danych

Pierwszy etap procesu ETL - Extract) - rozpoczyna się od napisania kodu w celu zebrania zestawów danych do dalszego sprawdzenia i przetworzenia. W tym celu przeskanujemy wszystkie foldery serwera produkcyjnego, zbierzymy dokumenty o określonym formacie i zawartości, a następnie przekonwertujemy je do postaci ustrukturyzowanej. Proces ten został szczegółowo omówiony w rozdziałach "Konwersja danych nieustrukturyzowanych i tekstowych do postaci ustrukturyzowanej" oraz "Konwersja danych CAD (BIM) do postaci ustrukturyzowanej" (Rys. 4.1-1 - Rys. 4.1-12).



Rys. 7.2-4 Przekształć dane CAD (BIM) w jedną dużą ramkę danych, która będzie zawierać wszystkie sekcje projektu.

Jako przykład ilustrujący, używamy kroku ładowania danych Extract i uzyskujemy tabelę wszystkich projektów CAD- (BIM-) (rys. 7.2-4) wykorzystuje konwertery z obsługą inżynierii odwrotnej [138] dla formatów RVT i IFC w celu uzyskania ustrukturyzowanych tabel ze wszystkich projektów i połączenia ich w jedną dużą tabelę DataFrame.



```

1 import os
2 import subprocess
3 import time
4 import pandas as pd
5
6 path_conv = r'C:\DDC_2023\' # Where RvtExporter.exe|IfcExporter.exe is located
7 path = r'C:\IFCprojects\' # Where Revit|IFC files are stored
8
9 def convert_and_wait(path_conv, exporter_name, file_path, extension):
10    # Start the conversion process
11    subprocess.Popen([os.path.join(path_conv, exporter_name),
12                     file_path], cwd=path_conv)
13    output_file = os.path.join(path,
14                               f"{os.path.splitext(os.path.basename(file_path))[0]}_{extension}.xlsx")
15
16    # Conversion process for RVT and IFC files
17    for file in os.listdir(path):
18        full_path = os.path.join(path, file)
19        if file.endswith('.ifc'):
20            convert_and_wait(path_conv, 'IfcExporter.exe', full_path, 'ifc')
21        elif file.endswith('.rvt'):
22            convert_and_wait(path_conv, 'RvtExporter.exe', full_path, 'rvt')
23
24    # Combine converted Excel files into one dataframe
25 df = pd.concat([pd.read_excel(os.path.join(path, f)) for f in os.listdir(path)
26 if f.endswith('.xlsx')], ignore_index=True)

```

Run in IDE

Rys. 7.2-5 Konwersja przy użyciu kodu Python i narzędzia inżynierii odwrotnej SDK dla plików RVT i IFC do jednej dużej struktury (df) DataFrame.

Pandas DataFrame może ładować dane z różnych źródeł, w tym plików tekstowych CSV, arkuszy kalkulacyjnych Excel, plików JSON - i XML -, formatów przechowywania dużych zbiorów danych, takich jak Parquet i HDF5, oraz z MySQL, PostgreSQL, SQLite, Microsoft SQL Server, Oracle i innych baz danych. Ponadto Pandas obsługuje ładowanie danych z interfejsów API, stron internetowych, usług w chmurze i systemów pamięci masowej, takich jak Google BigQuery, Amazon Redshift i Snowflake.

- 💡 Aby napisać kod do łączenia się i zbierania informacji z baz danych, wyślij podobne żądanie tekstowe do czatu LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN lub dowolnego innego):

Proszę o napisanie przykładu połączenia z MySQL i konwersji danych do ↵

💡 Odpowiedź LLM :

```

1 import pandas as pd
2 import mysql.connector
3
4 # Using a context manager to ensure
5 # that the connection is closed automatically
6 with mysql.connector.connect(host='hostname', database='dbname',
7     user='username', password='password') as conn:
8     df = pd.read_sql('SELECT * FROM tablename LIMIT 10', conn)

```

Create in LLM Chat

Run in IDE

Rys. 7.2-6 Przykład łączenia się przez Python z bazą danych MySQL i importowania danych z bazy danych MySQL do DataFrame .

Wynikowy kod (rys. 7.2-5, rys. 7.2-6) można uruchomić w jednym z popularnych IDE (zintegrowanych środowisk programistycznych), o których wspomnialiśmy powyżej w trybie offline: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse z wtyczką PyDev, Thonny, Wing IDE, IntelliJ IDEA z wtyczką Python, JupyterLab lub popularne narzędzia online: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Ładując dane wieloformatowe do zmiennej "df" (Rys. 7.2-5 - wiersz 25; Rys. 7.2-6 - wiersz 8), przekonwertowaliśmy dane do formatu Pandas DataFrame, jednej z najpopularniejszych struktur przetwarzania danych, która jest dwuwymiarową tabelą z wierszami i kolumnami. Więcej na temat innych formatów przechowywania danych wykorzystywanych w ETL -Pipelines, takich jak Parquet, Apache ORC, JSON, Feather, HDF5, a także nowoczesnych hurtowni danych, powiemy w rozdziale "Przechowywanie danych i zarządzanie nimi w branży budowlanej" (Rys. 8.1-2).

Po etapie ekstrakcji i strukturyzacji danych (Extract) powstaje pojedyncza tablica informacji (rys. 7.2-5, rys. 7.2-6), gotowa do dalszego przetwarzania. Jednak przed załadowaniem tych danych do systemów docelowych lub wykorzystaniem ich do analizy konieczne jest zapewnienie ich jakości, integralności i zgodności z określonymi wymaganiami. To właśnie na tym etapie odbywa się transformacja danych (Transform) - kluczowy krok zapewniający wiarygodność późniejszych wniosków i decyzji.

ETL Transform: zastosowanie reguł walidacji i transformacji

Krok Transform to etap, w którym dane są przetwarzane i przekształcane. Proces ten może obejmować

sprawdzanie poprawności, normalizację, uzupełnianie brakujących wartości i walidację przy użyciu zautomatyzowanych narzędzi

Według badania PwC "Data-Driven. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9], nowoczesne firmy audytorskie odchodzą od wyrywkowej weryfikacji danych i zmierzają w kierunku analizy ogromnych ilości informacji przy użyciu zautomatyzowanych narzędzi. Takie podejście umożliwia nie tylko identyfikację rozbieżności w raportowaniu, ale także oferowanie zaleceń dotyczących optymalizacji procesów biznesowych.

W budownictwie podobne metody mogą być stosowane na przykład do automatycznej walidacji danych projektowych, kontroli jakości konstrukcji i oceny wydajności wykonawcy. Jednym z narzędzi automatyzujących i przyspieszających przetwarzanie danych jest wykorzystanie wyrażeń regularnych (RegEx) w fazie transformacji danych (Transform) procesu ETL. RegEx pozwala efektywnie sprawdzać poprawność ciągów danych, wykrywać niespójności i zapewniać integralność informacji przy minimalnym zużyciu zasobów. Więcej szczegółów na temat RegEx (rys. 4.4-7) omówiono w rozdziale "Przekształcanie wymagań w ustrukturyzowaną formę".

Rozważmy praktyczny przykład: w systemie zarządzania obiektem nieruchomości (RPM) menedżer określa wymagania dla kluczowych atrybutów obiektów (rys. 7.2-7). Na etapie transformacji należy zweryfikować następujące parametry:

- weryfikacja formatów identyfikatorów obiektów (atribut "ID")
- kontrola wartości okresu gwarancji zastępczej (atribut "Okres gwarancji")
- weryfikacja cyklu wymiany elementu (atribut "Wymagania konserwacyjne")

Property Manager: Long-term Management

ID	Element	Warranty Period	Replacement Cycle	Maintenance Requirements
W-NEW	Window	-	20 years	Annual Inspection
W-OLD1	Window	8 years	15 years	Biannual Inspection
W-OLD2	Window	8 years	15 years	Biannual Inspection
D-122	Door	15 years	25 years	Biennial Maintenance

Rys. 7.2-7 Weryfikacja jakości rozpoczęta się od ustalenia wymagań dla atrybutów i ich wartości granicznych.

Aby ustawić wartości graniczne dla walidacji parametrów, założymy na przykład, że wiemy z naszego doświadczenia, że prawidłowe wartości dla atrybutu "ID" mogą zawierać tylko wartości łańcuchowe "W-NEW", "W-OLD1" lub "D-122" lub podobne wartości, w których pierwszym znakiem jest litera, po której następuje myślnik, a następnie trzy znaki alfabetyczne "NEW", "OLD" lub dowolna trzycyfrowa liczba (Rys. 7.2-7). Do weryfikacji tych identyfikatorów można użyć następującego wyrażenia regularnego (RegEx):

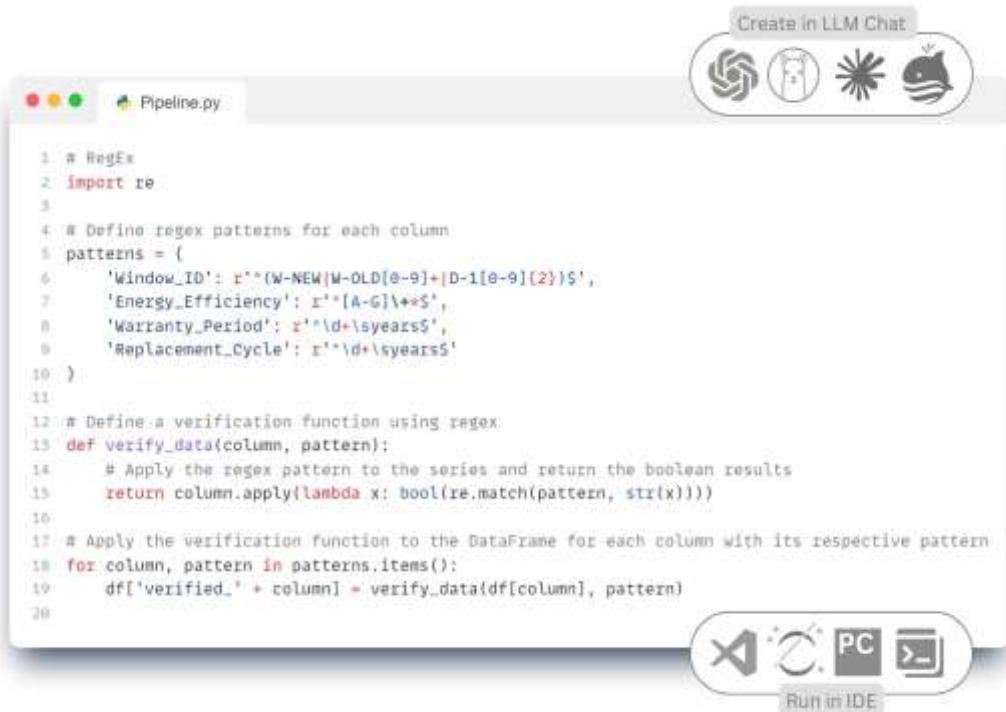
```
^W-NEW$| ^W-OLD[0-9]+$| ^D-1[0-9]{2}$
```

Ten szablon upewnia się, że wszystkie identyfikatory w danych spełniają określone kryteria. Jeśli jakakolwiek wartość nie przejdzie kontroli, system zarejestruje błąd. Aby utworzyć kod Python w celu przekształcenia danych i wykorzystania wynikowych danych do utworzenia tabeli wyników, wystarczy sformułować zapytanie na czacie LLM.

✉ Zgłoszenie tekstowe na adres LLM:

Napisz kod sprawdzający poprawność kolumn DataFrame za pomocą wyrażeń regularnych, który sprawdza identyfikatory w formacie "W-NEW" lub "W-OLD" za pomocą RegEx, efektywność energetyczną za pomocą liter od "A" do "G", okres gwarancji i cykl wymiany za pomocą wartości liczbowych w latach. ↵

✉ Odpowiedź LLM :



The screenshot shows a code editor window with Python code for validating DataFrame columns. The code uses regular expressions (RegEx) from the 're' library to define patterns for various columns: Window_ID, Energy_Efficiency, Warranty_Period, and Replacement_Cycle. It then defines a verification function 'verify_data' that applies these patterns to each column's series and returns boolean results. Finally, it iterates through the columns and applies the verification function to each, adding a new 'verified_' prefix to the resulting column names. The code editor interface includes tabs for 'Pipeline.py', a toolbar with icons for file operations, and a status bar at the bottom.

```
1 # RegEx
2 import re
3
4 # Define regex patterns for each column
5 patterns = {
6     'Window_ID': r"^(W-NEW|W-OLD[0-9]+|D-1[0-9]{2})$",
7     'Energy_Efficiency': r"^[A-G]\+*$",
8     'Warranty_Period': r"\d+\syears$",
9     'Replacement_Cycle': r"\d+\syears$"
10 }
11
12 # Define a verification function using regex.
13 def verify_data(column, pattern):
14     # Apply the regex pattern to the series and return the boolean results
15     return column.apply(lambda x: bool(re.match(pattern, str(x))))
16
17 # Apply the verification function to the DataFrame for each column with its respective pattern
18 for column, pattern in patterns.items():
19     df['verified_' + column] = verify_data(df[column], pattern)
20
```

Rys. 7.2-8 Kod automatyzuje proces walidacji poprzez zastosowanie szablonów RegEx do kolumn parametrów ramki danych .

Powyższy, automatycznie wygenerowany kod Pythona (Rys. 7.2-8), bibliotekę "re" (wyrażenia regularne RegEx) do zdefiniowania funkcji, która sprawdza każdy atrybut elementu danych w DataFrame. Dla każdej określonej kolumny (atrybutu) funkcja stosuje wzorzec RegEx, aby sprawdzić, czy każdy rekord jest zgodny z oczekiwany formatem i dodaje wyniki jako nowe wartości (False/True) do nowej kolumny atrybutu DataFrame.

Taka zautomatyzowana walidacja zapewnia formalną zgodność danych i może być wykorzystywana jako część systemu kontroli jakości na etapie transformacji.

Po pomyślnym zakończeniu etapu Transform i kontroli jakości, dane są gotowe do przesłania do systemów docelowych. Przekształcone i zweryfikowane dane można przesyłać do CSV, JSON, Excel, baz danych i innych formatów do dalszego wykorzystania. W zależności od zadania, wyniki mogą być również prezentowane w postaci raportów, wykresów lub pulpitów analitycznych.

ETL Load: Wizualizacja wyników na wykresach i grafach

Po zakończeniu etapu Transform, gdy dane zostały ustrukturyzowane i zweryfikowane, ostatnim etapem jest Load, w którym dane mogą być zarówno załadowane do systemu docelowego, jak i wizualizowane do analizy. Wizualna prezentacja danych pozwala szybko zidentyfikować odchylenia, przeanalizować rozkłady i przekazać kluczowe wnioski wszystkim uczestnikom projektu, w tym osobom bez wykształcenia technicznego.

Zamiast prezentować informacje w postaci tabel i liczb, możemy użyć infografik, wykresów i dashboardów (pulpitów nawigacyjnych). Jednym z najbardziej popularnych i elastycznych narzędzi do wizualizacji danych strukturalnych w Pythonie jest biblioteka Matplotlib (Rys. 7.2-9, Rys. 7.2-10). Umożliwia ona tworzenie statycznych, animowanych i interaktywnych wykresów oraz obsługuje szeroki zakres typów wykresów.

- Aby zwizualizować wyniki sprawdzania atrybutów z systemu RPM (Rys. 7.2-7), można użyć następującego zapytania do modelu językowego:

Napisz kod do wizualizacji danych DataFrame, powyżej (Rys. 7.2-7), z histogramem dla wyników, aby pokazać częstotliwość błędów w atrybucie ↴

- ❷ Odpowiedź LLM w formie kodu i gotowa wizualizacja wyników wykonania kodu bezpośrednio na czacie LLM :

The screenshot shows a software interface for generating code and visualizations. At the top, there's a toolbar with icons for 'Create in LLM Chat' and four decorative icons. Below that is a code editor window titled 'Pipeline.py' containing the following Python code:

```

1 # Re-importing necessary libraries for visualization
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4
5 # Visualization 1: Bar Chart
6 plt.figure(figsize=(10, 6))
7 df_visual.plot(kind='bar', stacked=True, color=['green', 'red'])
8 plt.title('Data Verification Summary - Bar Chart')
9 plt.xlabel('Data Categories')
10 plt.ylabel('Count')
11 plt.xticks(rotation=45)
12 plt.tight_layout()
13 plt.show()

```

At the bottom of the interface are buttons for 'Run in IDE' and other execution options. Below the interface is the generated visualization, a stacked bar chart titled 'Data Verification Summary - Bar Chart'. The chart shows the count of values checked for four categories: 'Window_ID', 'Energy_Efficiency', 'Warranty_Period', and 'Replacement_Cycle'. The y-axis represents the 'Count (of values checked)' from 0 to 5. The x-axis is labeled 'Data Categories'. A legend indicates that green represents 'Passed' and red represents 'Failed'. The data is summarized in the following table:

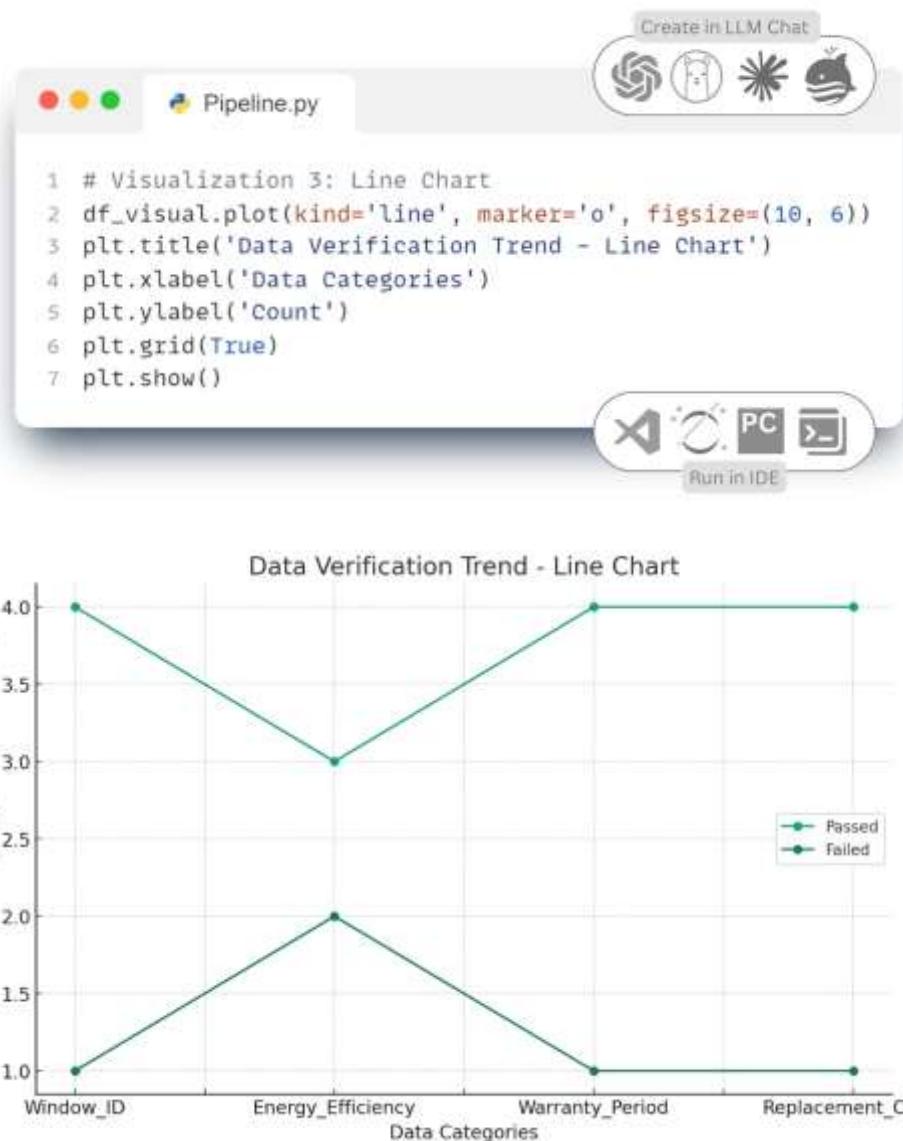
Data Category	Passed (Green)	Failed (Red)	Total Count
Window_ID	4	1	5
Energy_Efficiency	3	2	5
Warranty_Period	4	1	5
Replacement_Cycle	3	2	5

Rys. 7.2-9 Wizualizacja wyników kroku Transform sprawdzania wartości atrybutów z systemu RPM (rys. 7.2-7) jako histogram w kroku Load .

- Istnieje wiele darmowych bibliotek wizualizacyjnych typu open source, które umożliwiają prezentację ustrukturyzowanych danych w różnych formatach. Kontynuujmy wizualizację wyników za pomocą innego typu wykresu z następującym monitem na czacie:

Wykreś te same dane jako wykres liniowy ↴

- Odpowiedź LLM :



Rys. 7.2-10 Wizualizacja danych walidacyjnych (Rys. 7.2-8) jako wykresu liniowego uzyskanego przy użyciu biblioteki Matplotlib .

Istnieje wiele darmowych bibliotek wizualizacyjnych typu open source, takich jak

- Seaborn - dla wykresów statystycznych (rys. 7.2-11)
- Plotly - do interaktywnych wizualizacji internetowych (rys. 7.2-12, rys. 7.1-6)
- Altair - do wizualizacji deklaratywnej
- Dash lub Streamlit - do tworzenia pełnych pulpitów nawigacyjnych

Znajomość konkretnych bibliotek do wizualizacji nie jest konieczna - nowoczesne narzędzia, w tym LLM, pozwalają na automatyczne generowanie kodu wykresów i całych aplikacji na podstawie opisu zadania.

Wybór narzędzia zależy od celów projektu: czy jest to raport, prezentacja czy pulpit nawigacyjny online. Na przykład biblioteka open source Seaborn jest szczególnie dobra do pracy z danymi kategorycznymi, pomagając w identyfikacji wzorców i trendów.

- Aby zobaczyć działającą bibliotekę Seaborn, można bezpośrednio poprosić LLM o skorzystanie z wybranej biblioteki lub wysłać podobną prośbę tekstową do LLM na adres:

Pokaż mapę popularności dla ↪

- Odpowiedź LLM w postaci kodu i gotowego wykresu, którego kod wykreślania można teraz skopiować do IDE, a sam wykres można skopiować lub zapisać w celu wklejenia do dokumentu:

The screenshot shows a Streamlit application window titled "Pipeline.py". At the top right are icons for "Create in LLM Chat" and four AI models: a blue swirl icon, a white cat icon, a yellow starburst icon, and a blue whale icon. Below the title bar is a toolbar with icons for "X", "Run", "PC", and "Share". The main area contains Python code for generating a heatmap:

```

1 # Visualization 2: Heatmap
2 plt.figure(figsize=(8, 4))
3 sns.heatmap(df_visual, annot=True, cmap='viridis')
4 plt.title('Data Verification Summary - Heatmap')
5 plt.xlabel('Verification Status')
6 plt.ylabel('Data Categories')
7 plt.show()

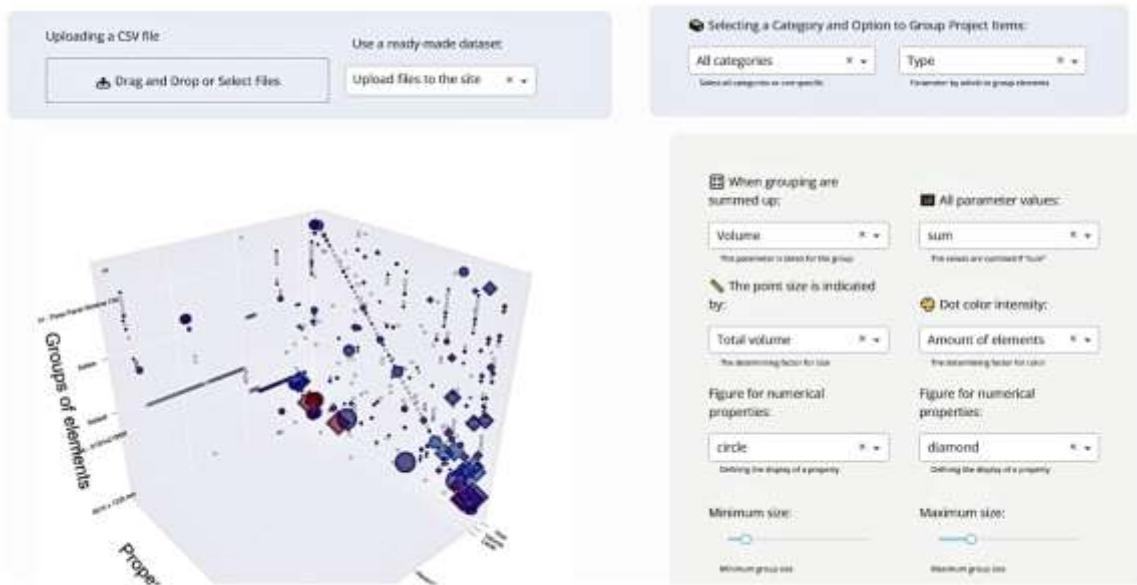
```

Below the code is a generated heatmap visualization with a color scale from red to green.



Rys. 7.2-11 Wizualizacja wyników walidacji (Rys. 7.2-8) danych przy użyciu biblioteki Seaborn.

Dla tych, którzy preferują podejście interaktywne, dostępne są narzędzia umożliwiające tworzenie dynamicznych wykresów i paneli z możliwością interakcji. Biblioteka Plotly (Rys. 7.1-6, Rys. 7.2-12) oferuje możliwość tworzenia wysoce interaktywnych wykresów i paneli, które mogą być osadzone na stronach internetowych i pozwalają użytkownikowi na interakcję z danymi w czasie rzeczywistym.



Rys. 7.2-12 Interaktywna wizualizacja 3D atrybutów elementów z projektu CAD- (BIM-) przy użyciu biblioteki Plotly.

Specjalistyczne biblioteki open source Bokeh, Dash i Streamlit zapewniają wygodny sposób prezentacji danych bez konieczności posiadania głębokiej wiedzy na temat tworzenia stron internetowych. Bokeh nadaje się do złożonych interaktywnych wykresów, Dash służy do tworzenia pełnoprawnych pulpitów analitycznych, a Streamlit pozwala szybko tworzyć aplikacje internetowe do analizy danych.

Dzięki takim narzędziom wizualizacyjnym programiści i analitycy mogą skutecznie rozpowszechniać wyniki

wśród współpracowników i interesariuszy, umożliwiając intuicyjną interakcję z danymi i upraszczając podejmowanie decyzji.

ETL Load: Automatyczne tworzenie dokumentów PDF

Na etapie ładowania danych można nie tylko wizualizować dane, przesyłać je do tabel lub baz danych, ale także automatycznie generować raporty, w tym niezbędne wykresy, diagramy i kluczowe wskaźniki analityczne, które otrzyma menedżer lub specjalista oczekujący na wyniki audytu. Zautomatyzowane raporty mogą zawierać zarówno komentarze i tekstową interpretację danych, jak i elementy wizualizacji - tabele, wykresy.

- Aby utworzyć raport PDF z histogramem (rys. 7.2-9) i opisem analizy opartej na walidacji, którą przeprowadziliśmy w poprzednich rozdziałach, wystarczy sformułować żądanie w kontynuacji dialogu z LLM, np.:

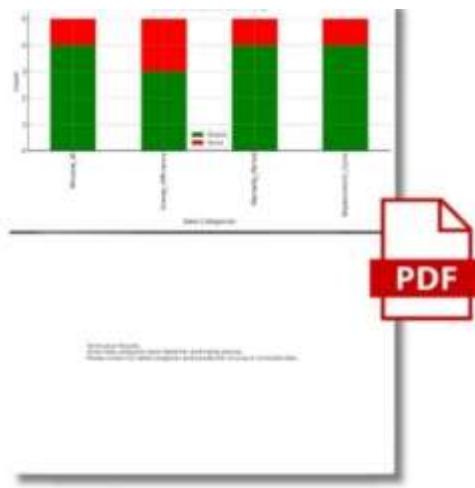
Napisz kod, aby utworzyć plik PDF z histogramem i opisem wyników powyższej walidacji danych (na czacie) oraz napisz ostrzeżenie tekstowe, że niektóre kategorie nie zostały zweryfikowane i że należy uzupełnić brakujące ↵

- Odpowiedź LLM jako kod i gotowy plik PDF z wynikami:

```

1 # Create a PDF document with the verification results
2 with PdfPages('verification_results.pdf') as pdf:
3     # Adding a histogram to the document
4     plt.figure()
5     df.pdf.plot(kind='bar', stacked=True, colour=['green', 'red'])
6     plt.title('Data Verification Summary')
7     plt.xlabel('Data Categories')
8     plt.ylabel('Count')
9     plt.tight_layout()
10    pdf.savefig()
11    plt.close()
12
13    # Adding a text page with a description
14    plt.figure()
15    plt.axis('off')
16    text_str = ("Verification Results:\n"
17                "Some data categories have failed the verification process.\n"
18                "Please review the failed categories and provide the missing data.")
19    plt.text(0.1, 0.5, text_str, wrap=True)
20    pdf.savefig()
21    plt.close()
22

```



Rys. 7.2-13 Zautomatyzowany kod tworzy dokument PDF zawierający histogram z danymi testowymi i tekstu z wynikami testu.

Automatycznie napisane rozwiązanie składające się zaledwie 20 linii kodu przy użyciu LLM natychmiast tworzy żądzany dokument PDF (lub DOC) z wizualizacją w postaci histogramu atrybutów (rys. 7.2-13) pokazującego liczbę danych, które przeszły i nie przeszły walidacji, oraz z dodatkiem bloku tekowego podsumowującego wyniki i zalecenia dalszych działań.

Zautomatyzowane generowanie dokumentów jest kluczowym elementem etapu Load, szczególnie w środowisku projektowym, w którym szybkość raportowania i dokładność mają kluczowe znaczenie.

ETL Load: automatyczne generowanie dokumentów z FPDF

Automatyzacja raportowania na etapie ETL Load jest ważnym krokiem w przetwarzaniu danych, zwłaszcza gdy wyniki analiz muszą być przedstawione w formacie, który jest łatwy do przekazania i zrozumienia. W branży budowlanej jest to często istotne w przypadku raportów z postępu prac, statystyk danych projektowych, raportów zapewnienia jakości lub dokumentacji finansowej.

Jednym z najwygodniejszych narzędzi do takich zadań jest biblioteka open source, FPDF, dostępna zarówno dla Pythona jak i PHP.

Biblioteka open source **FPDF** zapewnia elastyczny sposób generowania dokumentów za pomocą kodu, umożliwiając dodawanie nagłówków, tekstu, tabel i obrazów. Używanie kodu zamiast ręcznej edycji zmniejsza liczbę błędów i przyspiesza proces przygotowywania raportów w formacie PDF.

Jednym z kluczowych etapów tworzenia dokumentu PDF jest dodanie nagłówków i głównego tekstu w formie komentarzy lub opisów. Jednak podczas tworzenia raportu ważne jest nie tylko dodanie tekstu, ale także jego odpowiednia struktura. Nagłówki, wcięcia, odstępy między wierszami - wszystko to wpływa na czytelność dokumentu. Korzystając z FPDF, można ustawić parametry formatowania, kontrolować rozmieszczenie elementów i dostosować styl dokumentu.

FPDF jest bardzo podobny do HTML. Ci, którzy znają już HTML, mogą z łatwością generować dokumenty

PDF o dowolnej złożoności za pomocą FPDF, ponieważ struktura kodu jest bardzo podobna do znaczników HTML: nagłówki, tekst, obrazy i tabele są dodawane w podobny sposób. Ci, którzy nie są zaznajomieni z HTML, nie muszą się martwić - można użyć LLM, który natychmiast pomoże skompilować kod w celu wygenerowania pożądanego układu dokumentu.

- Poniższy przykład pokazuje, jak wygenerować raport z nagłówkiem i treścią. Wykonanie tego kodu w dowolnym IDE z obsługą Pythona tworzy plik PDF zawierający żądaną nagłówek i tekst:

```
from fpdf import FPDF      # Import biblioteki FPDF
pdf = FPDF()    # Utwórz PDF -document
pdf.add_page()   # Dodaj stronę

pdf.set_font("Arial", style='B', size=16) # Ustaw czcionkę: Arial, pogrubiona, rozmiar 16
pdf.cell(200, 10, "Raport projektu", ln=True, align='C') # Utwórz nagłówek i wyśrodkuj go
pdf.set_font("Arial", size=12) # Zmień czcionkę na zwykłą Arial, rozmiar 12
pdf.multi_cell(0, 10, "Ten dokument zawiera dane dotyczące wyników weryfikacji plików projektu...") # Dodaj tekst wielowierszowy
pdf.output(r "C:\reports\report.pdf") # Zapisz PDF -file
```



Rys. 7.2-14 Za pomocą kilku linijek kodu Pythona możemy automatycznie wygenerować potrzebny nam dokument tekstowy PDF.

Podczas przygotowywania raportów ważne jest, aby wziąć pod uwagę, że dane, z których tworzony jest dokument, rzadko są statyczne. Nagłówki, bloki tekstowe (rys. 7.2-14) są często tworzone dynamicznie, otrzymując wartości na etapie transformacji w procesie ETL.

Wykorzystanie kodu pozwala na tworzenie dokumentów zawierających aktualne informacje: nazwę projektu, datę wygenerowania raportu, a także informacje o uczestnikach czy aktualnym statusie. Wykorzystanie zmiennych w kodzie pozwala na automatyczne wstawianie tych danych w wymagane miejsca w raporcie, całkowicie eliminując konieczność ręcznej edycji przed wysłaniem.

Oprócz zwykłego tekstu i nagłówków, tabele zajmują szczególne miejsce w dokumentacji projektowej. Niemal każdy dokument zawiera ustrukturyzowane dane: od opisów obiektów po wyniki inspekcji. Automatyczne generowanie tabel na podstawie danych z etapu Transform pozwala nie tylko przyspieszyć proces przygotowania dokumentu, ale także zminimalizować błędy podczas przesyłania informacji. FPDF umożliwia wstawianie tabel do plików PDF (jako tekst lub obrazy), ustawianie obramowania komórek,

rozmiarów kolumn i czcionek (rys. 7.2-15). Jest to szczególnie wygodne podczas pracy z dynamicznymi danymi, gdy liczba wierszy i kolumn może się zmieniać w zależności od zadań dokumentu.

- Poniższy przykład pokazuje, jak zautomatyzować tworzenie tabel, np. z listami materiałów, szacunkami lub wynikami testów parametrów:

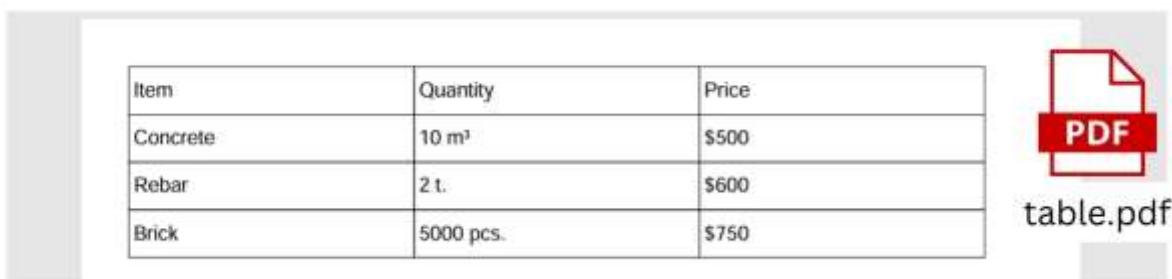
```

data = [
    [ "Pozycja", "Ilość", "Cena"], # Nagłówki kolumn
    [ "Beton", "10 m³", "$ 500." ], # Dane pierwszego wiersza
    [ "Rebar", "2 tonnes", "$ 600" ], # Dane drugiego wiersza .
    [ "Cegła", "5000 sztuk", "$ 750." ], # Dane wiersza 3.
]

pdf = FPDF () # Utwórz PDF -document
pdf.add_page() # Dodaj stronę
pdf.set_font("Arial", size=12) # Ustaw czcionkę

for row in data: # Przeszukuje wiersze tabeli
    for item in row: # Przejdz przez komórki w wierszu
        pdf.cell(60, 10, item, border=1) # Utwórz komórkę z obramowaniem, szerokością 60 i
wysokością 10
    pdf.ln() # Przejście do następnego wiersza
pdf.output(r "C:\reports\table.pdf") # Zapisz plik PDF -file

```



The screenshot displays a table with three columns: Item, Quantity, and Price. The table contains four rows of data: Concrete (10 m³, \$500), Rebar (2 t., \$600), and Brick (5000 pcs., \$750). To the right of the table is a red PDF icon with the text 'table.pdf' below it.

Item	Quantity	Price
Concrete	10 m³	\$500
Rebar	2 t.	\$600
Brick	5000 pcs.	\$750

Rys. 7.2-15 Można automatycznie generować nie tylko tekst, ale także dowolne informacje tabelaryczne z kroku Transform w PDF.

W rzeczywistych scenariuszach raportowania tabele są zwykle dynamicznie generowanymi informacjami uzyskanymi na etapie transformacji danych. W podanym przykładzie (rys. 7.2-15) tabela jest wstawiona do dokumentu PDF w formie statycznej: dane dla przykładu zostały umieszczone w słowniku danych (pierwsza linia kodu), w rzeczywistych warunkach taka zmienna danych jest wypełniana automatycznie po np. grupowaniu ramek danych.

W praktyce tabele takie budowane są często w oparciu o ustrukturyzowane dane pochodzące z różnych dynamicznych źródeł: baz danych, plików Excel, interfejsów API czy wyników obliczeń analitycznych. Najczęściej na etapie Transform (ETL) dane są agregowane, grupowane lub filtrowane - a dopiero potem przekształcane w sumy w postaci wykresów lub dwuwymiarowych tabel wyświetlanych w raportach. Oznacza to, że zawartość tabeli może się zmieniać w zależności od wybranych parametrów, okresu analizy, filtrów projektu lub ustawień użytkownika.

Wykorzystanie dynamicznych ramek danych i zestawów danych na etapie Transform sprawia, że proces raportowania na etapie Load jest tak elastyczny, skalowalny i łatwo powtarzalny, jak to tylko możliwe, bez konieczności ręcznej interwencji.

Oprócz tabel i tekstu FPDF obsługuje również dodawanie wykresów danych tabelarycznych, co pozwala na osadzanie w raporcie obrazów wygenerowanych za pomocą Matplotlib lub innych bibliotek wizualizacyjnych, które omówiliśmy powyżej. Dowolne wykresy, diagramy i diagramy można dodawać do dokumentu za pomocą kodu.

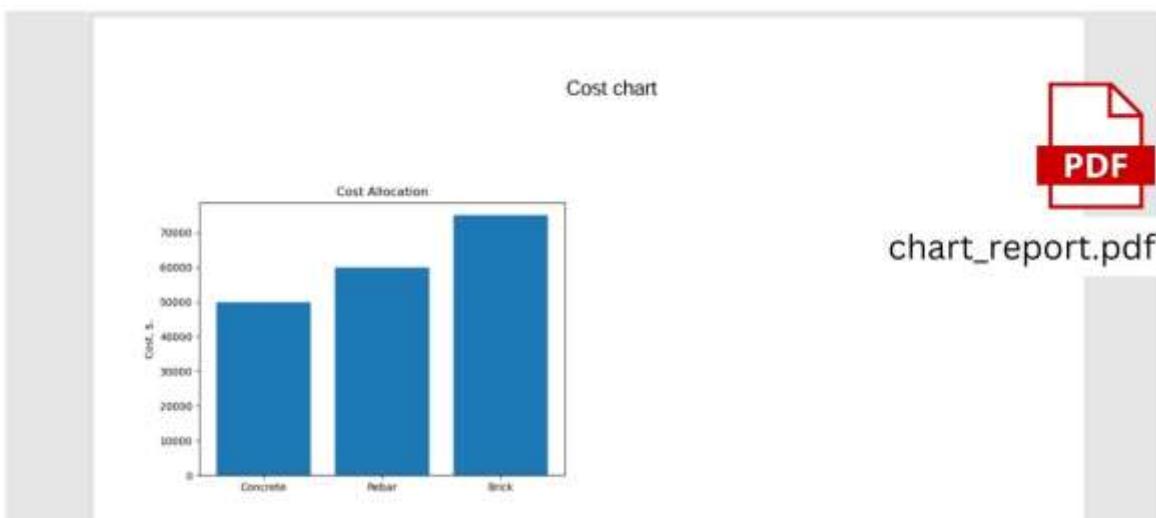
- Korzystając z biblioteki Pythona FPDF, dodajmy wykres wstępnie wygenerowany za pomocą Matplotlib. do dokumentu PDF:

```
import matplotlib.pyplot as plt # Import Matplotlib do tworzenia wykresów

fig, ax = plt.subplots() # Utwórz figurę i osie wykresu
categories = ["Concrete", "Rebar", "Brick"] # Nazwy kategorii
values = [50000, 60000, 75000] # Wartości kategorii
ax.bar(categories, values) # Utwórz wykres słupkowy
plt.ylabel("Value,$.") # Podpisz os Y
plt.title("Cost Distribution") # Dodaj tytuł
plt.savefig(r "C:\reports\chart\chart\chart.png") # Zapisz wykres jako obrazek

pdf = FPDF () # Utwórz PDF -document
pdf.add_page() # Dodaj stronę
pdf.set_font("Arial", size=12) # Ustaw czcionkę
pdf.cell(200, 10, "Cost Chart", ln=True, align='C') # Dodaj nagłówek

pdf.image(r "C:\reports\chart\chart\chart.png", x=10, y=30, w=100) # Wstaw obraz
do PDF (x, y - współrzędne, w - szerokość)
pdf.output(r "C:\reports\chart_report.pdf") # Zapisz plik PDF
```



Rys. 7.2-16 Za pomocą kilkunastu linijek kodu można wygenerować wykres, zapisać go, a następnie wkleić do dokumentu PDF.

FPDF czyni proces przygotowania dokumentów przejrzystym, szybkim i wygodnym. Wbudowane w kod szablony pozwalają na generowanie dokumentów z aktualnymi danymi, eliminując konieczność ich ręcznego wypełniania.

Korzystanie z automatyzacji ETL - zamiast czasochłonnego ręcznego raportowania, specjalści mogą skupić się na analizie danych i podejmowaniu decyzji, zamiast wybierać odpowiednie narzędzie do pracy z konkretnym silosem danych z przejrzystym interfejsem użytkownika.

W ten sposób biblioteka FPDF zapewnia elastyczne narzędzie do automatycznego tworzenia dokumentów o dowolnej złożoności - od krótkich raportów technicznych po złożone podsumowania analityczne z tabelami i wykresami, co pozwala nie tylko przyspieszyć obieg dokumentów, ale także znacznie zmniejszyć prawdopodobieństwo błędów związanych z ręcznym wprowadzaniem danych i formatowaniem.

ETL Load: Raportowanie i ładowanie do innych systemów

Na etapie Load wyniki zostały wygenerowane w postaci tabel, wykresów i końcowych raportów PDF przygotowanych zgodnie z ustalonymi wymaganiami. Ponadto możliwe jest wyeksportowanie tych danych do formatów nadających się do odczytu maszynowego (np. CSV), co jest niezbędne do integracji z zewnętrznymi systemami, takimi jak ERP, CAFM, CPM, platformy BI i inne rozwiązania korporacyjne lub branżowe. Oprócz CSV, dane mogą być przesyłane do XLSX, JSON, XML lub bezpośrednio do baz danych, które obsługują automatyczną wymianę informacji.

- 💡 Aby wygenerować odpowiedni kod w celu zautomatyzowania kroku Load, wystarczy wysłać zapytanie do interfejsu LLM, na przykład: ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude lub QWEN:

Napisz kod, aby wygenerować raport wyników walidacji danych w DataFrame, gdzie kolumny z prefiksem "verified_" są zliczane, zmieniane na "Passed" i "Failed", brakujące wartości są zastępowane 0, a następnie tylko te wiersze, które przeszły wszystkie walidacje, są eksportowane do pliku CSV.

💡 Odpowiedź LLM:



```

1 # Create a report on the results of data validation
2 report = df.filter(like='verified_').apply(lambda x: x.value_counts()).T
3 report.columns = ['Passed', 'Failed']
4 report.fillna(0, inplace=True)
5
6 # Export the cleaned and verified data to a CSV file
7 verified_df = df[df.filter(like='verified_').all(axis=1)]
8 verified_df.to_csv('verified_data.csv', index=False)

```

The screenshot shows a code editor window titled "Pipeline.py". The code itself is as follows:

```

1 # Create a report on the results of data validation
2 report = df.filter(like='verified_').apply(lambda x: x.value_counts()).T
3 report.columns = ['Passed', 'Failed']
4 report.fillna(0, inplace=True)
5
6 # Export the cleaned and verified data to a CSV file
7 verified_df = df[df.filter(like='verified_').all(axis=1)]
8 verified_df.to_csv('verified_data.csv', index=False)

```

At the top of the window, there are icons for "Create in LLM Chat" and four AI models. At the bottom, there are icons for "Run in IDE", "Run in Jupyter", "Run in Colab", and "Run in PC".

Rys. 7.2-17 Zweryfikowane dane uzyskane w kroku Transform z końcowej ramki danych są eksportowane do pliku CSV w celu integracji z innymi systemami.

Powyższy kod (rys. 7.2-17) implementuje ostatni etap procesu ETL - Load, podczas którego zweryfikowane dane są zapisywane w formacie CSV, kompatybilnym z większością zewnętrznych systemów i baz danych. W ten sposób zakończyliśmy pełny cykl procesu ETL, w tym ekstrakcję, transformację, wizualizację, dokumentację i eksport danych do potrzebnych nam systemów i formatów, co zapewnia powtarzalność, przejrzystość i automatyzację pracy z informacjami.

Potok ETL może być wykorzystywany zarówno do przetwarzania pojedynczych projektów, jak i do zastosowań na dużą skalę - podczas analizy setek lub tysięcy przychodzących danych w postaci dokumentów, obrazów, skanów, projektów CAD, chmur punktów, plików PDF lub innych źródeł pochodzących z systemów rozproszonych. Zdolność do pełnej automatyzacji procesu sprawia, że ETL nie jest tylko technicznym narzędziem przetwarzania, ale podstawą cyfrowej infrastruktury informacji budowlanej.

ETL z LLM: Wizualizacja danych z dokumentów PDF

Nadszedł czas, aby przejść do budowy pełnoprawnego procesu ETL, który obejmuje wszystkie kluczowe etapy obsługi danych w jednym scenariuszu - ekstrakcję, transformację i ładowanie. Zbudujmy zautomatyzowany ETL-Pipeline, który pozwala przetwarzać dokumenty PDF bez pracy ręcznej - wyodrębnić dane z dokumentów, wizualizować je, analizować i przesyłać do innych systemów.

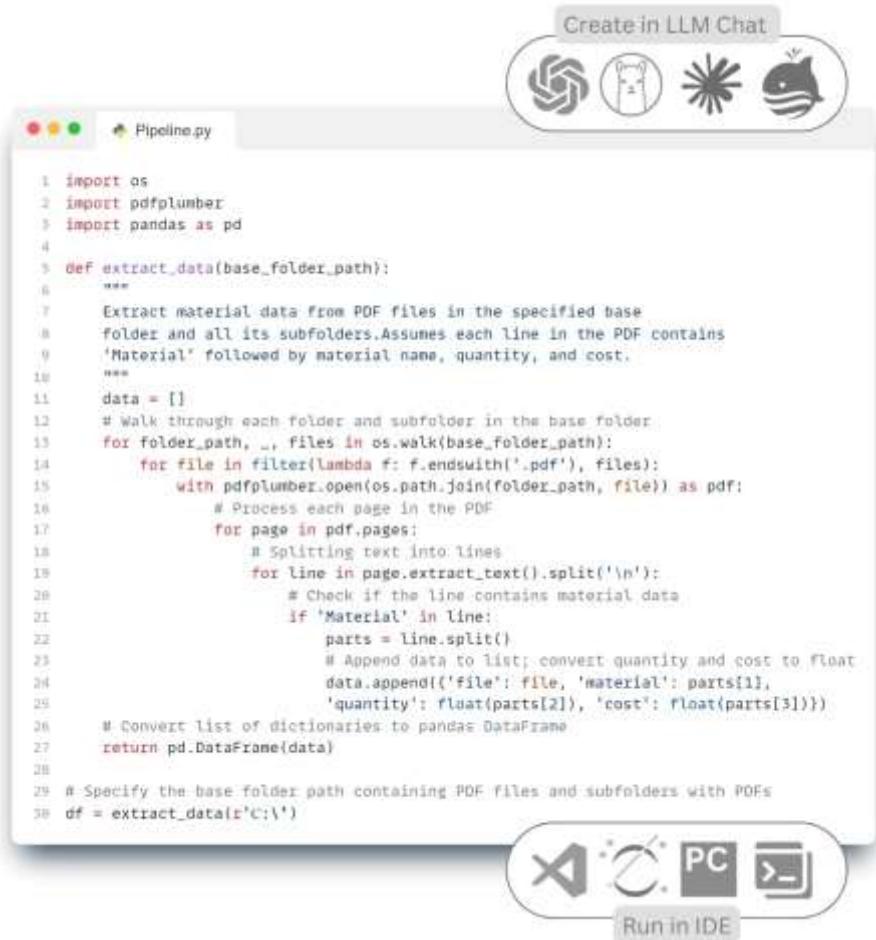
Proces ETL w naszym przykładzie zostanie opisany za pomocą podpowiedzi, które będą musiały wyjaśnić

modelowi językowemu (LLM) wszystkie procesy ETL wraz z opisem końcowego wyniku, który ma zostać uzyskany. W tym przypadku zadaniem jest znalezienie wszystkich plików PDF w określonym folderze i jego podfolderach, wyodrębnienie z nich odpowiednich informacji - np. nazw materiałów, ilości i kosztów - oraz przedstawienie wyniku w postaci ustrukturyzowanej tabeli (DataFrame) do dalszej analizy

- 💡 Pierwsze zapytanie tekstowe w LLM w celu automatycznego wyodrębnienia danych z wielu dokumentów PDF i utworzenia ramki danych dla kroku Wyodrębnij:

Napisz kod, aby wyodrębnić informacje o materiałach z plików PDF w danym folderze i jego podfolderach. Dane w pliku PDF obejmują nazwę materiału, ilość i koszt. Wynik powinien zostać zapisany w DataFrame ↴

- 💡 Odpowiedź LLM :



The screenshot shows a window titled "Pipeline.py" containing Python code. At the top right, there are icons for "Create in LLM Chat", "Run in IDE", and "PC". Below the code, there is another set of icons for "Run in IDE", "PC", and "VS Code". A "Run in IDE" button is also present at the bottom.

```

1 import os
2 import pdfplumber
3 import pandas as pd
4
5 def extract_data(base_folder_path):
6     """
7         Extract material data from PDF files in the specified base
8         folder and all its subfolders. Assumes each line in the PDF contains
9         'Material' followed by material name, quantity, and cost.
10    """
11    data = []
12    # Walk through each folder and subfolder in the base folder
13    for folder_path, _, files in os.walk(base_folder_path):
14        for file in filter(lambda f: f.endswith('.pdf'), files):
15            with pdfplumber.open(os.path.join(folder_path, file)) as pdf:
16                # Process each page in the PDF
17                for page in pdf.pages:
18                    # Splitting text into lines
19                    for line in page.extract_text().split('\n'):
20                        # Check if the line contains material data
21                        if 'Material' in line:
22                            parts = line.split()
23                            # Append data to list; convert quantity and cost to float
24                            data.append({'file': file, 'material': parts[1],
25                                      'quantity': float(parts[2]), 'cost': float(parts[3])})
26    # Convert list of dictionaries to pandas DataFrame
27    return pd.DataFrame(data)
28
29 # Specify the base folder path containing PDF files and subfolders with PDFs
30 df = extract_data(r'C:\V')

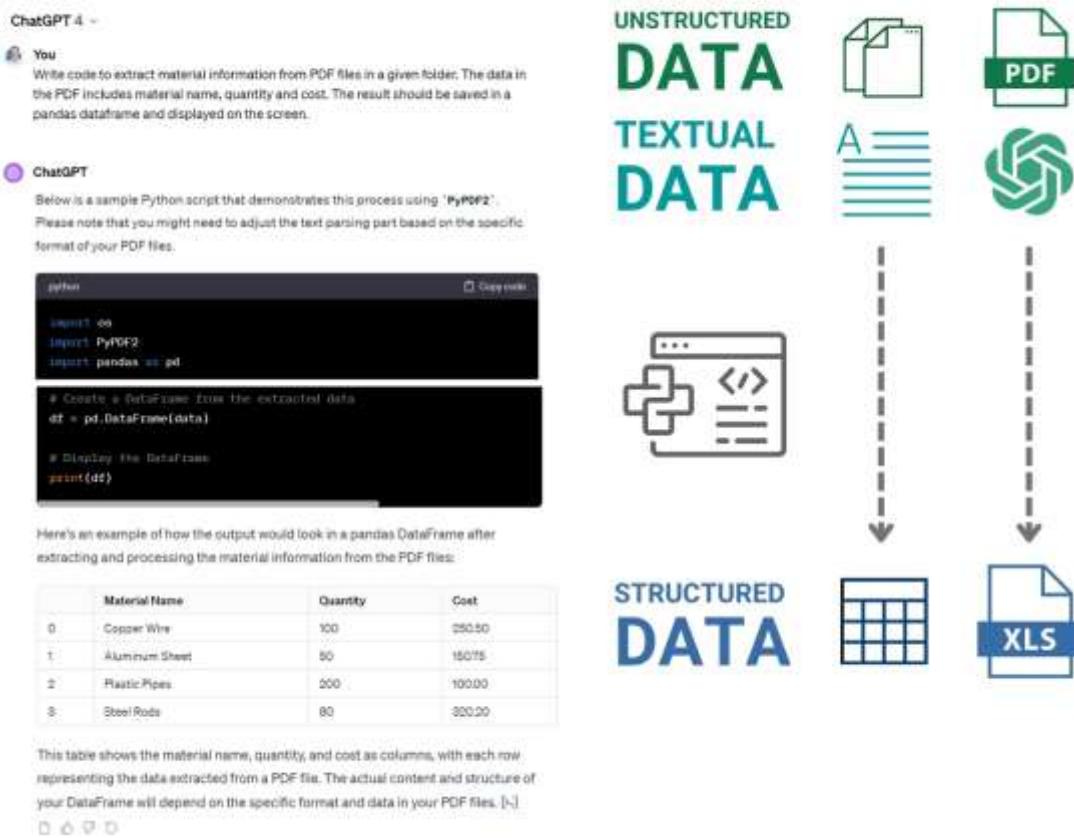
```

Rys. 7.2-18 LLM tworzy kod w języku Python do wyodrębniania danych z plików PDF w określonym folderze i wszystkich jego podfolderach.

Odpowiedź LLM (Rys. 7.2-18) to gotowy skrypt Pythona, który automatycznie przeszukuje wszystkie foldery, otwiera znalezione pliki PDF, wyodrębnia z nich informacje tekstowe i konwertuje je na tabelę. Wynikowy

kod czatu można uruchomić w jednym z popularnych IDE PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse z wtyczką PyDev, Thonny, Wing IDE, IntelliJ IDEA z wtyczką Python, JupyterLab lub popularnych narzędziach internetowych Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Alternatywnie, aby uprościć proces zbierania, zamiast kopiować kod z LLM i używać kodu w IDE, możemy również przesłać tuzin plików PDF bezpośrednio do czatu LLM (Rys. 7.2-19) i uzyskać tabelę jako dane wyjściowe, bez konieczności oglądania kodu lub uruchamiania go. Wynikiem wykonania tego kodu będzie tabela z wybranymi przez nas atrybutami.



Rys. 7.2-19 Wynik wykonania kodu w LLM, który wyodrębnia dane z plików PDF w ustrukturyzowanym widoku ramki danych z wybranymi atrybutami.

W kolejnym kroku prosimy o stworzenie modelu językowego na uzyskanych danych - na przykład w celu porównania kosztów i ilości zużycia materiałów oraz stworzenia przykładowych wizualizacji, które posłużą jako podstawa do dalszej analizy.

- Poproś na czacie z LLM o wykreślenie kilku wykresów z tabel utworzonych w kroku Przekształcanie (Rys. 7.2-18):

Wizualizacja całkowitego kosztu i ilości każdego materiału z DataFrame (Rys. 7.2-18)

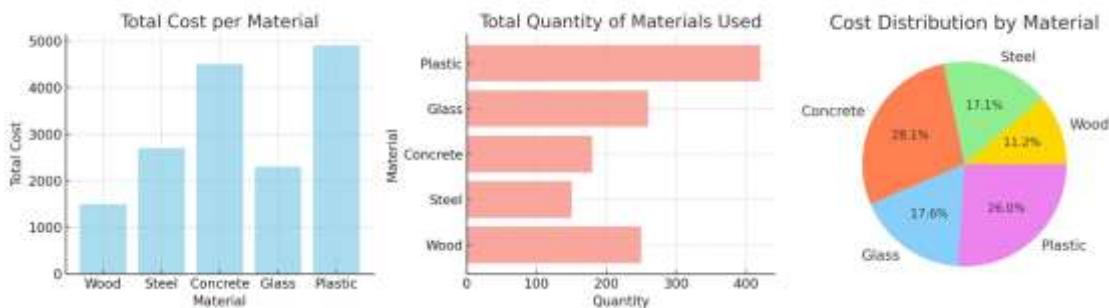
```

1 import matplotlib.pyplot as plt
2
3 # Plotting total cost per material
4 # Group data by 'material' and sum up the 'cost' for each material
5 df.groupby('material')['cost'].sum().plot(kind='bar')
6 plt.title('Total Cost per Material')
7 plt.xlabel('Material')
8 plt.ylabel('Total Cost')
9 plt.show()
10
11 # Plotting total quantity of materials used
12 # Group data by 'material' and sum up the 'quantity' for each material
13 df.groupby('material')['quantity'].sum().plot(kind='barh')
14 plt.title('Total Quantity of Materials Used')
15 plt.xlabel('Quantity')
16 plt.ylabel('Material')
17 plt.show()
18

```

Rys. 7.2-20 LLM response -models as Python code do wizualizacji danych z ramki danych przy użyciu biblioteki matplotlib.

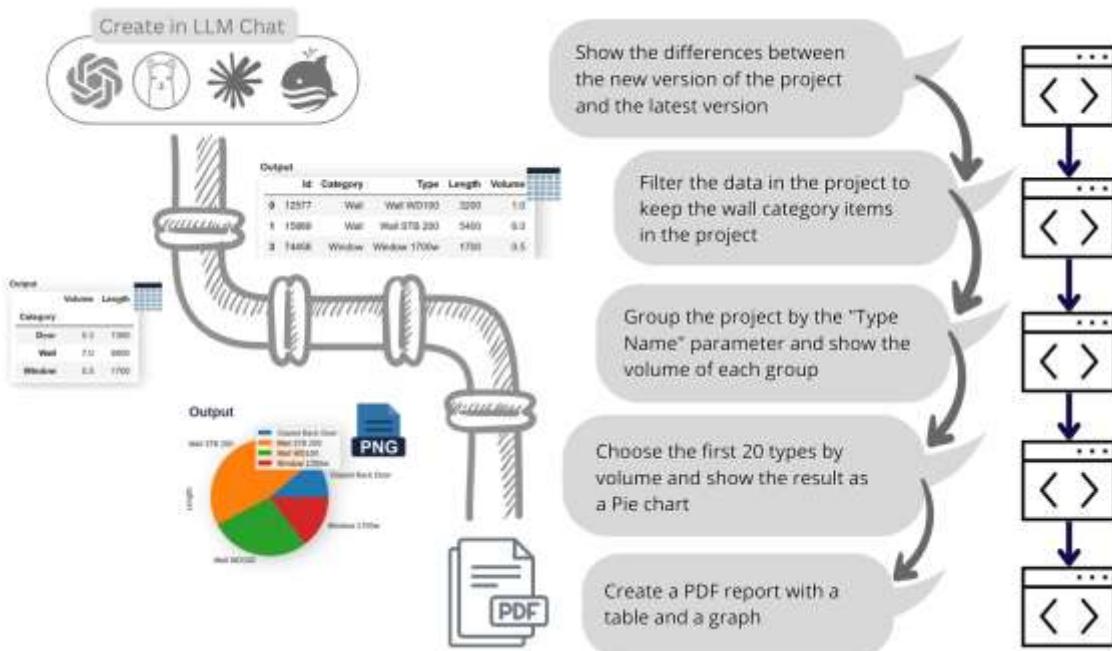
LLM automatycznie generuje i wykonuje kod Python (rys. 7.2-20) przy użyciu biblioteki matplotlib. Po wykonaniu tego kodu otrzymujemy wykresy kosztów i zużycia materiałów w projektach budowlanych bezpośrednio na czacie (rys. 7.2-21), co znacznie upraszcza pracę analityczną.



Rys. 7.2-21 Wizualizacja odpowiedzi LLM jako wykresy oparte na danych zebranych w DataFrame.

Wsparcie w opracowywaniu pomysłów na pisanie kodu ETL, analizowanie i wykonywanie kodu oraz wizualizowanie wyników jest dostępne za pośrednictwem prostych zapytań tekstowych w LLM, bez konieczności uczenia się podstaw programowania. Pojawienie się narzędzi AI, takich jak LLM, zdecydowanie zmienia podejście do programowania i automatyzacji przetwarzania danych (Rys. 7.2-22).

Według raportu PwC "Jaka jest prawdziwa wartość sztucznej inteligencji dla Twojej firmy i jak możesz ją wykorzystać?" (2017) [139], automatyzacja procesów i poprawa produktywności będą głównymi czynnikami napędzającymi wzrost gospodarczy. Oczekuje się, że poprawa produktywności będzie odpowiadać za ponad 55% całego wzrostu PKB napędzanego przez sztuczną inteligencję w latach 2017-2030".



Rys. 7.2-22 AI LLM pomaga wygenerować wersję roboczą kodu, która jest stosowana w przyszłych projektach bez potrzeby LLM.

Korzystając z narzędzi takich jak ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok, a także otwartych danych i oprogramowania open source, możemy zautomatyzować procesy, które wcześniej były wykonywane tylko za pomocą wyspecjalizowanych, kosztownych i trudnych w utrzymaniu modułowych systemów własnościowych.

W kontekście budownictwa oznacza to, że firmy, które jako pierwsze wdrożą zautomatyzowane procesy Pipeline -data, odniosą znaczące korzyści, począwszy od poprawy wydajności zarządzania projektami, poprzez zmniejszenie strat finansowych, aż po eliminację rozdrobnionych aplikacji i silosowych hurtowni danych.

Opisana logika wykonywania zadań biznesowych w procesie ETL jest kluczową częścią automatyzacji procesów analitycznych i przetwarzania danych, która jest specyficzną odmianą szerszego pojęcia - rurociągów (Pipelines).

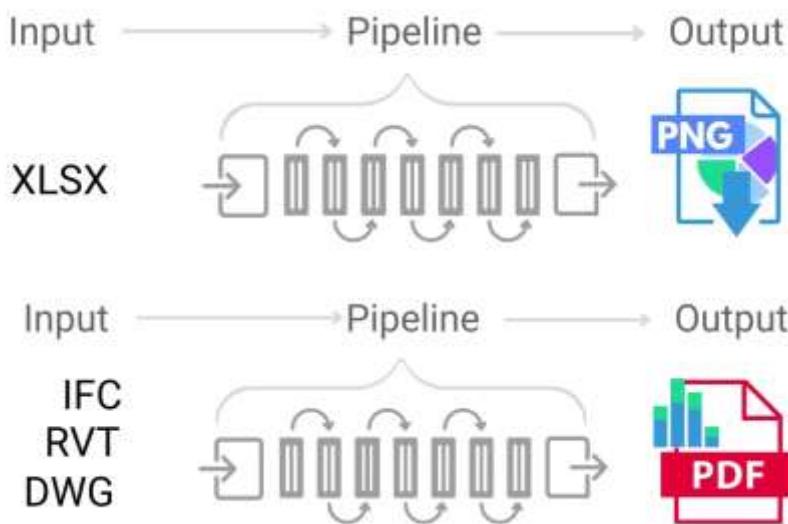
ROZDZIAŁ 6.3.

AUTOMATYCZNY PRZENOŚNIK ETL (RUROCIĄG)

Pipeline: Automatyczny przenośnik ETL danych

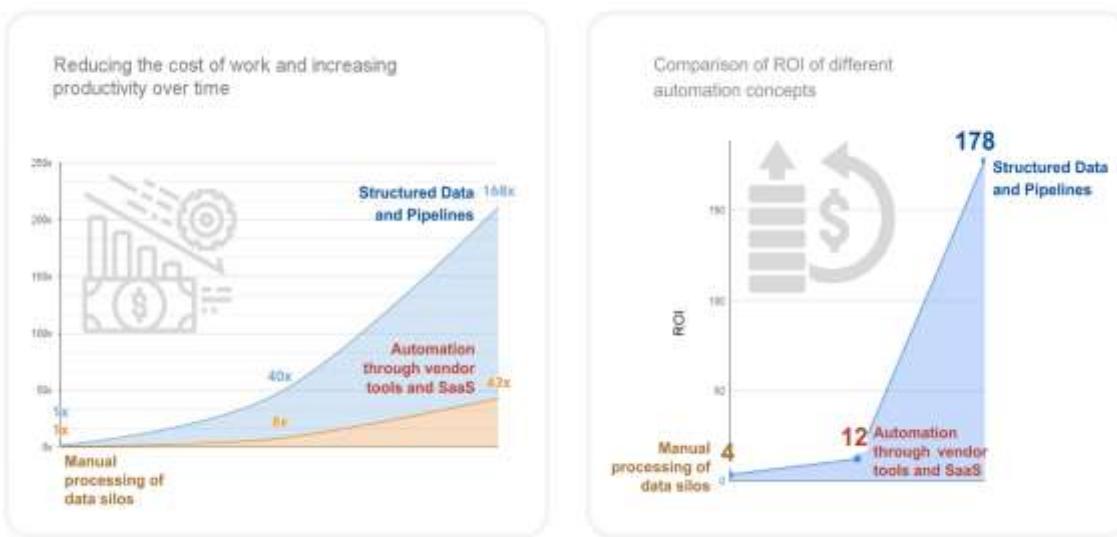
Proces ETL był tradycyjnie używany do przetwarzania danych w systemach analitycznych, obejmujących zarówno źródła ustrukturyzowane, jak i nieustrukturyzowane. Jednak w dzisiejszym środowisku cyfrowym coraz częściej używa się szerszego terminu - Pipeline (przenośnik), który opisuje dowolny sekwencyjny łańcuch przetwarzania, w którym dane wyjściowe jednego etapu stają się danymi wejściowymi dla następnego.

Podejście to dotyczy nie tylko danych, ale także innych rodzajów automatyzacji: przetwarzania zadań, tworzenia raportów, integracji z oprogramowaniem i cyfrowego przepływu pracy (rys. 7.3-1).



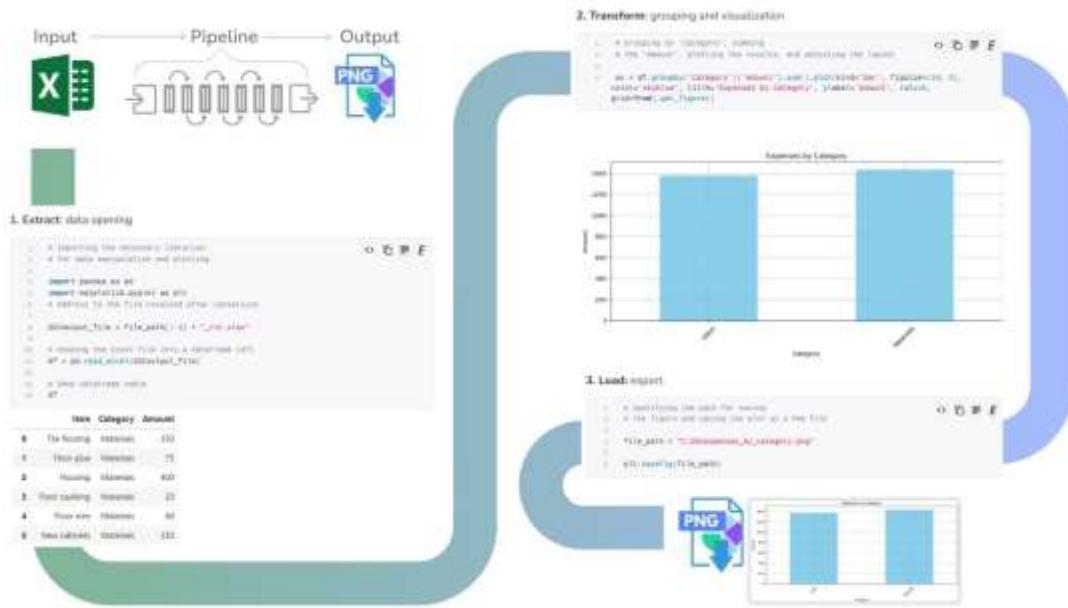
Rys. 7.3-1 Pipeline to sekwencja przetwarzania, w której wyjście jednego etapu staje się wejściem do następnego etapu.

Wykorzystanie Pipeline jest jednym z głównych elementów automatyzacji, szczególnie podczas pracy z dużymi ilościami heterogenicznych danych. Architektura Pipeline pozwala na organizację złożonych etapów przetwarzania w modułowym, spójnym i łatwym do zarządzania formacie, co zwiększa czytelność, upraszcza konserwację kodu i umożliwia przyrostowe debugowanie i skalowalne testowanie.



Rys. 7.3-2 ROI Proces walidacji danych Pipeline skraca czas wykonania o dziesiątki i setki razy w porównaniu z przetwarzaniem przy użyciu klasycznych narzędzi [74].

W przeciwieństwie do pracy ręcznej w zastrzeżonych systemach (ERP, PMIS, CAD, itp.), pipelining pozwala znacznie (rys. 7.3-2) zwiększyć szybkość wykonywania zadań, uniknąć powtarzalnej pracy i zautomatyzować uruchamianie procesów we właściwym czasie (rys. 7.3-3).

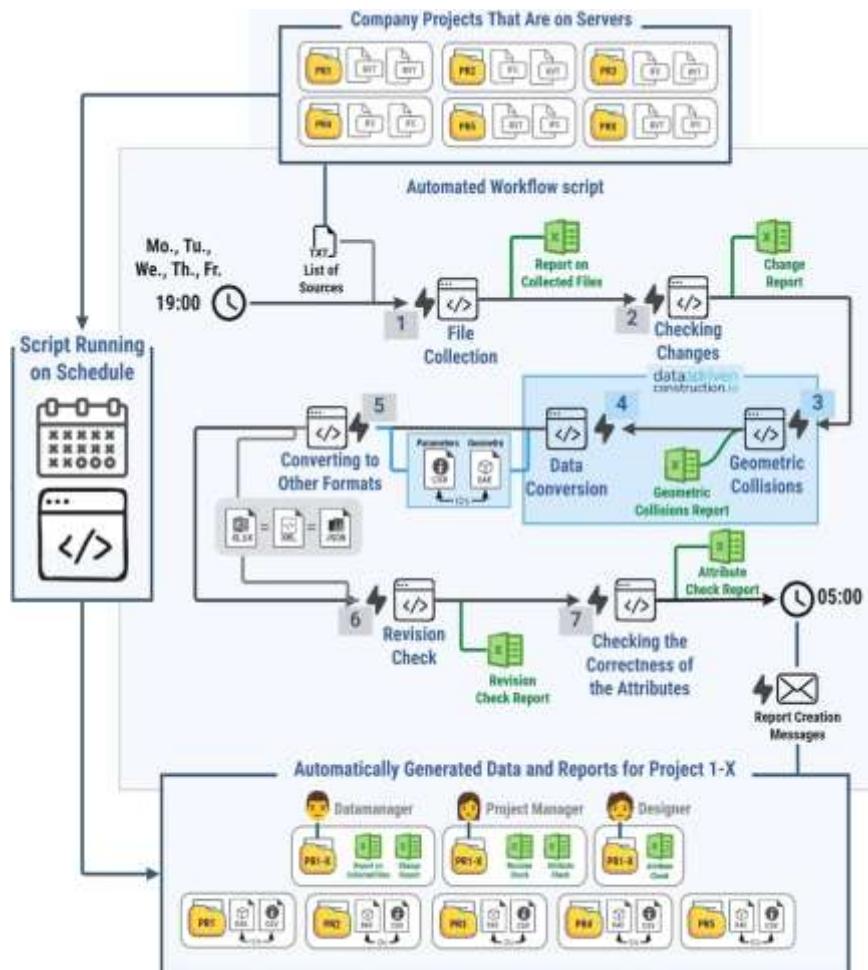


Rys. 7.3-3 Przykład ETL Potok do automatycznego tworzenia wykresu z danych tabelarycznych w pliku XLSX bez otwierania programu Excel.

Aby przetwarzać dane strumieniowe i zbudować zautomatyzowany Pipeline, podobny do procesu ETL, należy wcześniej zidentyfikować źródła danych, a także ramy czasowe ich gromadzenia - albo dla konkretnego procesu biznesowego, albo dla całej firmy.

W projektach budowlanych dane pochodzą z wielu heterogenicznych źródeł o różnej częstotliwości aktualizacji. Aby stworzyć wiarygodną wizytówkę danych, kluczowe jest rejestrowanie, kiedy informacje są pobierane i aktualizowane. Umożliwia to podejmowanie decyzji na czas i poprawia efektywność zarządzania projektem.

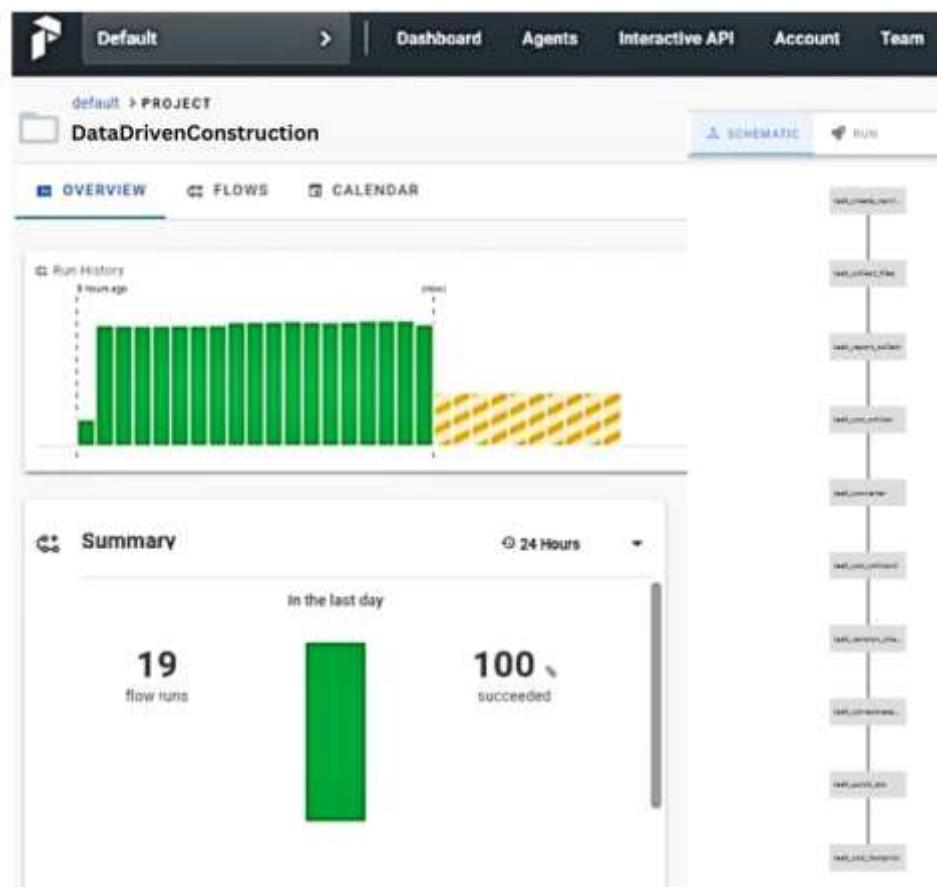
Jedną z opcji jest rozpoczęcie procesu montażu o ustalonej godzinie - na przykład o 19:00, na koniec dnia roboczego. W tym momencie uruchamiany jest pierwszy skrypt odpowiedzialny za agregację danych z różnych systemów i magazynów (rys. 7.3-4 krok 1). Następnie następuje zautomatyzowane przetwarzanie i transformacja danych do ustrukturyzowanego formatu odpowiedniego dla analityki (rys. 7.3-4 krok 2-4). Na ostatnim etapie, przy użyciu przygotowanych danych, automatycznie generowane są raporty, pulpity nawigacyjne i inne produkty opisane w poprzednich rozdziałach (rys. 7.3-4 krok 6-7). W rezultacie do godziny 05:00 rano menedżerowie mają już aktualne raporty o stanie projektu w wymaganym formacie (rys. 7.3-5).



Rys. 7.3-4 Dane w Pipeline, automatycznie zbierane wieczorem, są przetwarzane przez noc, dzięki czemu rano menedżerowie mają do dyspozycji aktualne i świeże raporty.

Terminowe gromadzenie danych, definiowanie KPI, automatyzacja procesów transformacji i wizualizacja za pomocą pulpitów nawigacyjnych to kluczowe elementy skutecznego podejmowania decyzji w oparciu o dane.

Takie zautomatyzowane procesy (rys. 7.3-4) mogą być wykonywane z pełną autonomią: działają zgodnie z harmonogramem, przetwarzają dane bez nadzoru i mogą być wdrażane w chmurze lub na własnym serwerze firmy (rys. 7.3-5). Pozwala to na integrację takich potoków ETL z istniejącą infrastrukturą IT, zachowując kontrolę nad danymi i zapewniając elastyczność w skalowaniu.



Rys. 7.3-5 Automatyczne procesy ETL -conveyor (rys. 7.3-4) na platformie Prefect, w których 10 skryptów Python jest uruchamianych naprzemiennie po godzinie 19:00 każdego dnia roboczego.

Automatyzacja przepływów pracy nie tylko zwiększa produktywność zespołu, zwalniając czas na bardziej znaczące i mniej rutynowe zadania, ale także służy jako ważny pierwszy krok w kierunkułączenia technologii sztucznej inteligencji (AI) do procesów biznesowych, które omówimy bardziej szczegółowo w rozdziale poświęconym analizie predykcyjnej i uczeniu maszynowemu.

Pipeline -ETL proces walidacji danych z LLM

W poprzednich rozdziałach dotyczących tworzenia wymagań dotyczących danych i automatyzacji ETL, krok po kroku podzieliliśmy proces przygotowania danych, transformacji, walidacji i wizualizacji. Działania te zostały zaimplementowane jako oddzielne bloki kodu (Rys. 7.2-18 - Rys. 7.2-20), z których każdy wykonuje określone zadanie.

Teraz mamy kolejny cel - połączenie tych elementów w jeden, spójny i zautomatyzowany potok przetwarzania danych - potok, ETL -Pipeline - w którym wszystkie etapy (ładowanie, walidacja, wizualizacja, eksport) są wykonywane sekwencyjnie w jednym automatycznie wykonywanym skrypcie.

W poniższym przykładzie zaimplementowany zostanie kompletny cykl przetwarzania danych: od załadowania źródłowego pliku CSV → do sprawdzenia struktury i wartości za pomocą wyrażeń regularnych → obliczanie wyników → generowanie raportu wizualnego w formacie PDF.

- 💡 Możesz użyć następującego zapytania tekstowego do LLM, aby pobrać odpowiedni kod:

Proszę o napisanie przykładowego kodu, który ładuje dane z CSV, waliduje DataFrame za pomocą wyrażeń regularnych, sprawdza identyfikatory w formacie "W-NEW" lub "W-OLD", efektywność energetyczną za pomocą liter od "A" do "G", okres gwarancji i cykl wymiany za pomocą wartości liczbowych w latach, a na koniec tworzy raport z liczbą wartości pozytywnych i negatywnych, generuje plik PDF z histogramem wyników i dodaje opis tekstowy. ↴

■ Odpowiedź LLM :

```

1 # Importing necessary libraries
2 import pandas as pd
3 import re
4 import matplotlib.pyplot as plt
5 from matplotlib.backends.backend_pdf import PdfPages
6
7
8 # Load data
9 df = pd.read_csv('path_to_csv_file.csv')
10
11 # Define patterns and verify data
12 patterns = {
13     'Window_ID': r'^W-NEW|W-OLD[0-9]+|D-1[0-9]{2}$',
14     'Energy_Efficiency': r'^[A-G]\+*$',
15     'Warranty_Period': r'^\d+\syears$',
16     'Replacement_Cycle': r'^\d+\syears$'
17 }
18
19 for column, pattern in patterns.items():
20     df[f'verified_{column}'] = df[column].apply(lambda x: bool(re.match(pattern, str(x))))
21
22 # Create report
23 report = df.filter(like='verified_').apply(lambda x: x.value_counts()).T
24 report.columns = ['Passed', 'Failed']
25
26 # Generate PDF
27 pdf_path = '/mnt/data/verification_results_full_pipeline.pdf'
28 with PdfPages(pdf_path) as pdf:
29     # Plot and save bar chart
30     report.plot(kind='bar', stacked=True, color=['green', 'red'])
31     plt.title('Data Verification Summary')
32     plt.xlabel('Data Categories')
33     plt.ylabel('Count')
34     plt.tight_layout()
35     pdf.savefig()
36     plt.close()
37
38 # Add text page
39 plt.figure()
40 plt.axis('off')
41 text_str = "Verification Results Summary"
42 plt.text(0.1, 0.5, text_str, wrap=True)
43 pdf.savefig()
44 plt.close()
45
46 pdf_path
47

```

Rys. 7.3-6 Pipeline (ETL) automatyzuje pełny cykl przetwarzania danych: od pobierania i weryfikacji do tworzenia ustrukturyzowanego raportu w formacie PDF .

Zautomatyzowany kod (rys. 7.3-6) wewnątrz czatu LLM lub w DIE, po skopiowaniu kodu, zweryfikuje dane z pliku CSV przy użyciu określonych wyrażeń regularnych, utworzy raport o liczbie pozytywnych i negatywnych rekordów, a następnie zapisze wyniki walidacji jako plik PDF.

Ta struktura ETL -conveyor, w której każdy krok - od ładowania danych do generowania raportów - jest zaimplementowany jako oddzielny moduł, zapewnia przejrzystość, skalowalność i powtarzalność. Przedstawienie logiki walidacji jako łatwego do odczytania kodu Python sprawia, że proces jest przejrzysty i zrozumiały nie tylko dla programistów, ale także dla specjalistów w zakresie zarządzania danymi, jakości i analityki.

Podejście Pipeline do automatyzacji przetwarzania danych pozwala na standaryzację procesów, zwiększenie ich powtarzalności i uproszczenie adaptacji do nowych projektów. Tworzy to ujednoliczoną metodologię analizy danych, niezależnie od źródła lub rodzaju zadania - niezależnie od tego, czy jest to testowanie zgodności, raportowanie czy przesyłanie danych do systemów zewnętrznych.

Taka automatyzacja ogranicza błędy ludzkie, zmniejsza zależność od zastrzeżonych rozwiązań oraz zwiększa dokładność i wiarygodność wyników, dzięki czemu nadaje się zarówno do analityki operacyjnej na poziomie projektu, jak i analityki strategicznej na poziomie firmy.

Rurociąg -ETL: weryfikacja danych i informacji o elementach projektu w CAD (BIM)

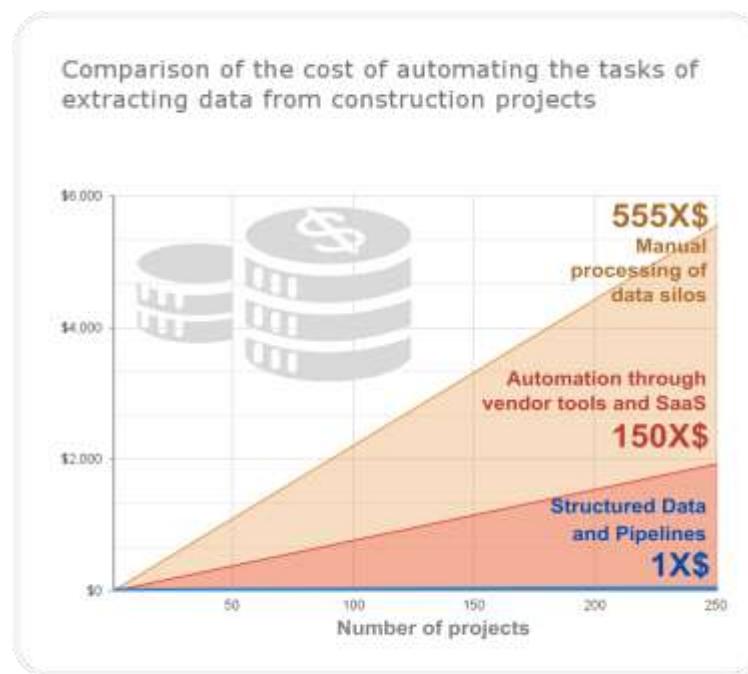
Dane z systemów CAD i baz danych (BIM) są jednymi z najbardziej zaawansowanych i dynamicznie aktualizowanych źródeł danych w działalności firm budowlanych. Aplikacje te nie tylko opisują projekt za pomocą geometrii, ale także uzupełniają go o wiele warstw informacji tekstowych: objętości, właściwości materiałów, przypisania pomieszczeń, poziomy efektywności energetycznej, tolerancje, cykle życia i inne atrybuty.

Atrybuty przypisane do podmiotów w modelach CAD są tworzone na etapie projektowania i stają się podstawą dalszych procesów biznesowych, w tym kalkulacji kosztów, harmonogramowania, oceny cyklu życia i integracji z systemami ERP i CAFM, gdzie wydajność procesów w dużej mierze zależy od jakości danych pochodzących z działów projektowych.

Tradycyjne podejście do walidacji atrybutów w modelach CAD- (BIM-) obejmuje ręczną walidację (rys. 7.2-1), która staje się długim i kosztownym procesem, gdy ilość modeli jest duża. Biorąc pod uwagę ilość i liczbę nowoczesnych projektów budowlanych oraz ich regularne aktualizacje, proces walidacji i transformacji danych staje się niezrównoważony i nieopłacalny.

Generalni wykonawcy i kierownicy projektów stają przed koniecznością przetwarzania dużych ilości danych projektowych, w tym wielu wersji i fragmentów tych samych modeli. Dane pochodzą od organizacji projektowych w formatach RVT, DWG, DGN, IFC, NWD i innych (Rys. 3.1-14) i wymagają regularnego sprawdzania pod kątem zgodności ze standardami branżowymi i korporacyjnymi

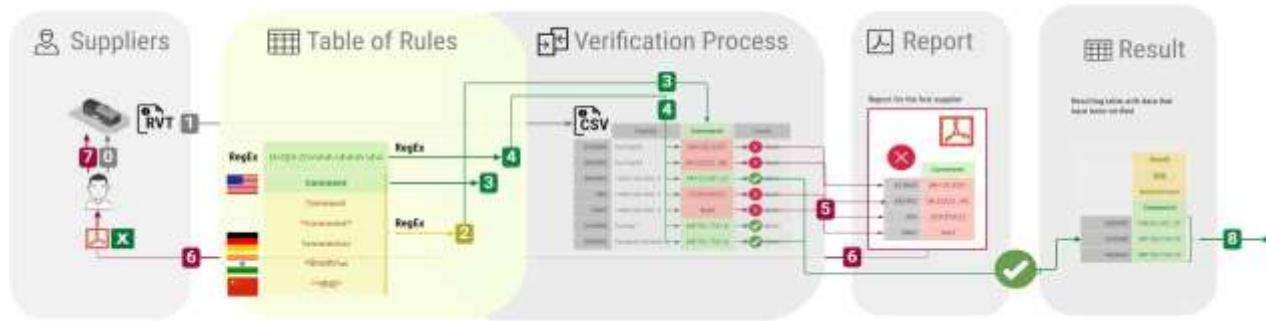
Zależność od ręcznych działań i specjalistycznego oprogramowania sprawia, że proces walidacji danych jest wąskim gardłem w przepływach pracy związanych z danymi z modeli obejmujących całą firmę. Automatyzacja i wykorzystanie ustrukturyzowanych wymagań może wyeliminować tę zależność, znacznie zwiększając szybkość i niezawodność walidacji danych (Rys. 7.3-7).



Rys. 7.3-7 Automatyzacja zwiększa szybkość weryfikacji i przetwarzania danych, co dziesiątki razy zmniejsza koszty pracy [140].

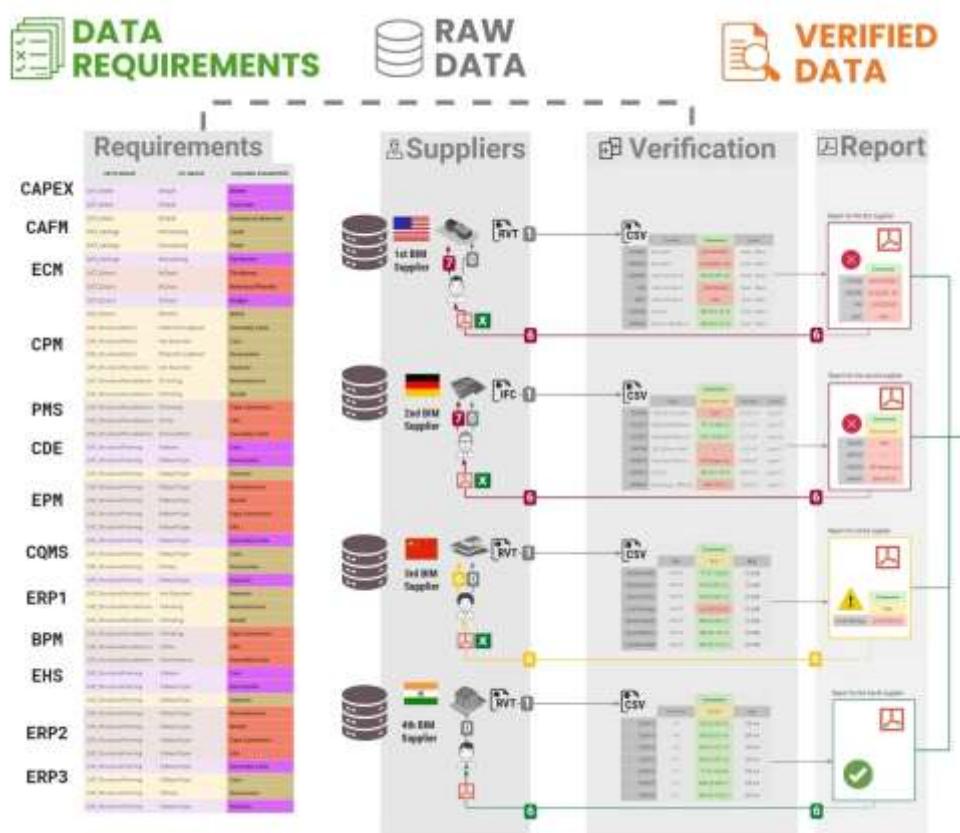
Proces walidacji danych CAD obejmuje ekstrakcję danych (etap ETL Extract) z różnych zamkniętych (RVT, DWG, DGN, NWS itp.) lub otwartych formatów półstrukturalnych i parametrycznych (IFC, CPXML, USD.) lub otwartych formatów półstrukturalnych i parametrycznych (IFC, CPXML, USD), w których tabele reguł mogą być stosowane do każdego atrybutu i jego wartości (etap Transform) przy użyciu wyrażeń regularnych RegEx (rys. 7.3-8), proces, który omówiliśmy szczegółowo w czwartej części książki.

Tworzenie raportu o błędach w formacie i pomyślnie zweryfikowanych rekordów powinno być zakończone danymi wyjściowymi (krok Load) w ustrukturyzowanych formatach, które uwzględniają tylko zweryfikowane podmioty, które można wykorzystać w dalszych procesach.



Rys. 7.3-8 Proces walidacji danych od dostawców danych projektu do raportu końcowego walidowany za pomocą wyrażeń regularnych.

Automatyzacja walidacji danych z systemów CAD (BIM) z ustrukturyzowanymi wymaganiami i przesyłanie strumieniowe nowych danych, które są przetwarzane przez ETL-Pipelines (Rys. 7.3-9) zmniejsza potrzebę ręcznego zaangażowania w proces walidacji (każdy z procesów walidacji i wymagań dotyczących danych został omówiony w poprzednich rozdziałach).



Rys. 7.3-9 Automatyzacja walidacji danych poprzez ETL upraszcza zarządzanie projektem budowlanym poprzez przyspieszenie procesów.

Tradycyjnie walidacja modeli dostarczanych przez wykonawców i specjalistów CAD (BIM) może trwać od

kilku dni do kilku tygodni. Jednak dzięki wprowadzeniu zautomatyzowanych procesów ETL, czas ten można skrócić do kilku minut. W typowej sytuacji wykonawca stwierdza: "*Model jest zwalidowany i zgodny*". To stwierdzenie rozpoczęyna łańcuch weryfikacji wykonawcy dotyczącego jakości danych:

- ☛ Kierownik projektu - "*Wykonawca twierdzi, że model został przetestowany, wszystko jest w porządku*".
- ☛ Menedżer danych - Weryfikacja obciążenia:
 - Prosty skrypt w Pandas wykrywa naruszenie w ciągu kilku sekund. Automatyzacja eliminuje spory:
 - Kategoria: OST_StructuralColumns, Parametr: FireRating IS NULL.
 - Wygenerowanie listy identyfikatorów naruszeń → eksport do programu Excel/PDF.

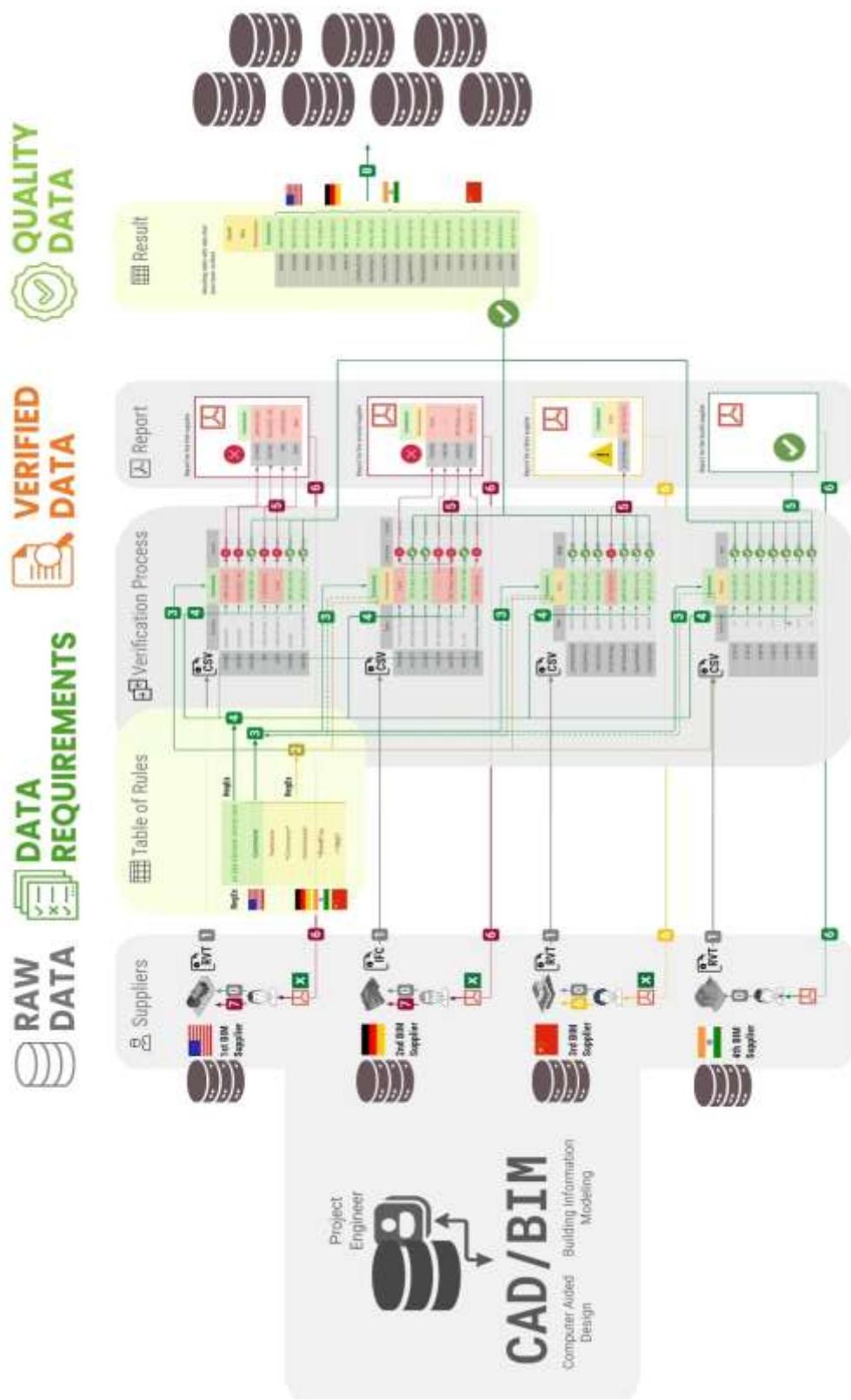
Prosty skrypt w Pandas wykrywa naruszenie w ciągu kilku sekund:

```
df = model_data[model_data["Category"] == "OST_StructuralColumns"] # Filtrowanie
issues = df[df["FireRating"].isnull()] # Puste wartości
issues[["ElementID"]].to_excel("fire_rating_issues.xlsx") # Eksport identyfikatorów
```

- ☛ Menedżer danych do kierownika projektu - "*Sprawdzenie strony wykazało, że 18 kolumn nie ma wypełnionego parametru FireRating*".
- ☛ Kierownik projektu do wykonawcy - "*Model został zwrócony do rewizji: parametr FireRating jest obowiązkowy, bez niego akceptacja jest niemożliwa*".

W rezultacie model CAD nie podlega walidacji, automatyzacja eliminuje spory, a wykonawca niemal natychmiast otrzymuje ustrukturyzowany raport z listą identyfikatorów problematycznych elementów. W ten sposób proces walidacji staje się przejrzysty, powtarzalny i chroniony przed błędami ludzkimi (rys. 7.3-10).

Takie podejście sprawia, że proces weryfikacji danych staje się funkcją inżynierijną, a nie ręczną kontrolą jakości. Nie tylko zwiększa to produktywność, ale także umożliwia zastosowanie tej samej logiki do wszystkich projektów firmy, umożliwiając kompleksową cyfrową transformację procesów, od projektowania po operacje.



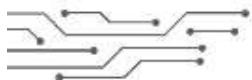
Rys. 7.3-10 Automatyzacja sprawdzania atrybutów elementów eliminuje błędy ludzkie i zmniejsza prawdopodobieństwo ich wystąpienia.

Dzięki wykorzystaniu zautomatyzowanych potoków (rys. 7.3-10), użytkownicy systemu oczekujący wysokiej jakości danych z systemów CAD- (BIM-) mogą natychmiast uzyskać potrzebne dane wyjściowe - tabele, dokumenty, obrazy - i szybko zintegrować je z zadaniami roboczymi.

Automatyzacja kontroli, przetwarzania i analizy napędza zmiany w podejściu do zarządzania projektami budowlanymi, w szczególności interoperacyjność różnych systemów, bez użycia złożonych i kosztownych modułowych, zastrzeżonych systemów lub zamkniętych rozwiązań dostawców.

Podczas gdy koncepcje i akronimy marketingowe pojawiają się i znikają, same procesy walidacji wymagań dotyczących danych na zawsze pozostaną integralną częścią procesów biznesowych. Zamiast tworzyć coraz bardziej wyspecjalizowane formaty i standardy, branża budowlana powinna sięgnąć po narzędzia, które dowiodły już swojej skuteczności w innych branżach. Obecnie istnieją potężne platformy do automatyzacji przetwarzania danych i integracji procesów, które pozwalają firmom znacznie skrócić czas poświęcany na rutynowe operacje i zminimalizować błędy w Extract, Transform i Load.

Jednym z popularnych przykładów rozwiązań do automatyzacji i orkiestracji procesów ETL jest Apache Airflow, który pozwala organizować złożone procesy obliczeniowe i zarządzać potokami ETL. Oprócz Airflow, aktywnie wykorzystywane są również inne podobne rozwiązania, takie jak Apache NiFi do routingu i przesyłania strumieniowego danych oraz n8n do automatyzacji procesów biznesowych.



ROZDZIAŁ 6.4.

ORKIESTRACJA ETL I PRZEPŁYWÓW PRACY: PRAKTYCZNE ROZWIĄZANIA

DAG i Apache Airflow: automatyzacja i orkiestracja przepływu pracy

Apache Airflow to darmowa i otwarta platforma, zaprojektowana do automatyzacji, orkiestracji i monitorowania przepływów pracy (ETL -conveyors).

Praca z dużymi ilościami danych jest wymagana każdego dnia:

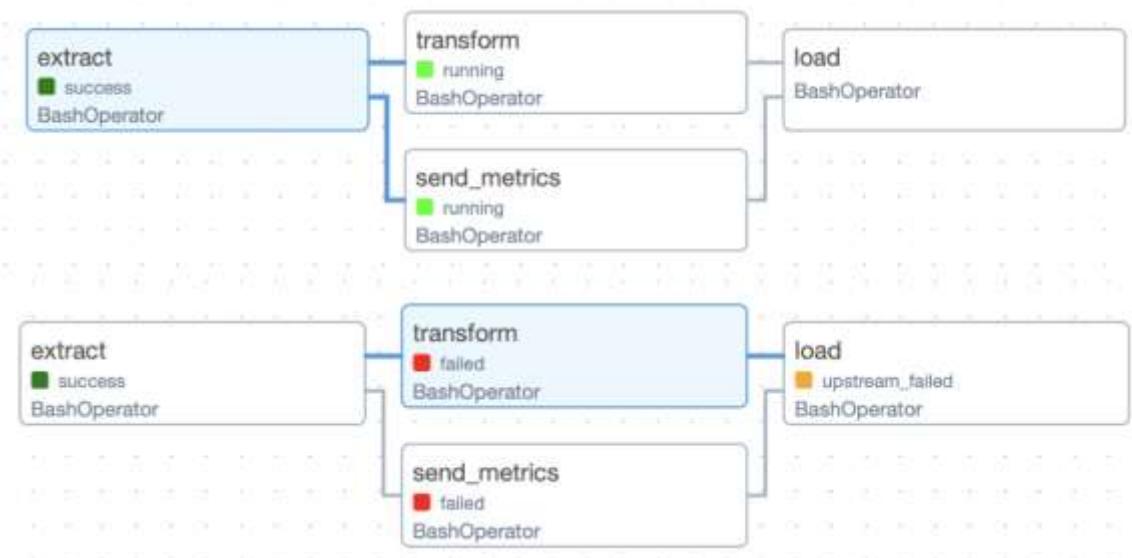
- Pobieranie plików z różnych źródeł - Wyodrębnianie (na przykład od dostawców lub klientów).
- Przekształcenie tych danych do wymaganego formatu - Przekształcenie (struktura, czyszczenie i weryfikacja)
- Wysyłanie wyników do weryfikacji i tworzenie raportów - Ładowanie (przesyłanie do wymaganych systemów, dokumentów, baz danych lub pulpitów nawigacyjnych).

Ręczne wykonywanie takich procesów ETL zajmuje dużo czasu i prowadzi do ryzyka błędu ludzkiego. Zmiana źródła danych lub awaria na jednym z etapów może spowodować opóźnienia i nieprawidłowe wyniki.

Narzędzia do automatyzacji, takie jak Apache Airflow, pozwalają zbudować niezawodny system ETL - conveyor, zminimalizować błędy, skrócić czas przetwarzania i zapewnić poprawność danych na każdym etapie. Sercem Apache Airflow jest koncepcja DAG (Directed Acyclic Graph) - skierowany graf acykliczny, w którym każde zadanie (operator) jest połączone z innymi zależnościami i wykonywane ściśle w określonej kolejności. DAG eliminuje cykle, co zapewnia logiczną i przewidywalną strukturę wykonywania zadań.

Airflow dba o orkiestrację - zarządzanie zależnościami między zadaniami, kontrolowanie harmonogramów wykonania, śledzenie statusu i automatyczne reagowanie na awarie. Takie podejście minimalizuje ręczną interwencję i zapewnia niezawodność całego procesu.

Task Orchestrator to narzędzie lub system zaprojektowany do zarządzania i kontrolowania wykonywania zadań w złożonych środowiskach obliczeniowych i informatycznych. Ułatwia proces wdrażania, automatyzacji i zarządzania wykonywaniem zadań w celu poprawy wydajności i optymalizacji zasobów.



Rys. 7.4-1 Apache Airflow zapewnia przyjazny dla użytkownika interfejs, w którym można wizualizować DAG -ETL, przeglądać dzienniki wykonania, stan uruchomienia zadania i nie tylko.

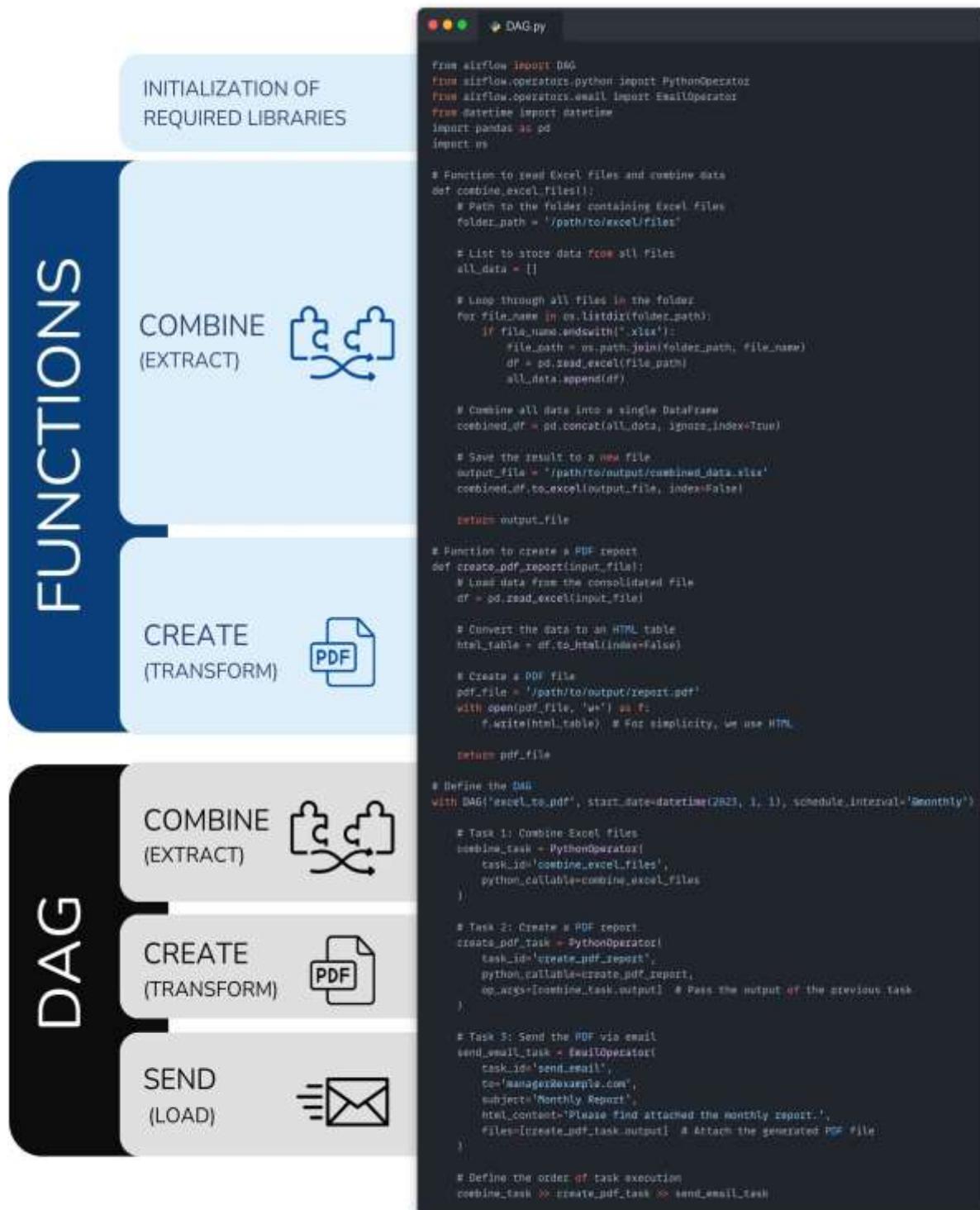
Airflow jest szeroko stosowany do orkiestracji i automatyzacji obliczeń rozproszonych, przetwarzania danych, zarządzania procesami ETL (Extract, Transform, Load), planowania zadań i innych scenariuszy danych. Domyślnie Apache Airflow używa SQLite jako bazy danych.

Przykład prostego grafu DAG, podobnego do ETL, składa się z zadań - Extract, Transform i Load. W grafie, który jest kontrolowany za pomocą interfejsu użytkownika (rys. 7.4-1), zdefiniowana jest kolejność wykonywania zadań (fragmentów kodu): na przykład najpierw wykonywany jest ekstrakt, następnie transformacja (i wysyłanie metryk), a zadanie ładowania kończy pracę. Gdy wszystkie zadania zostaną zakończone, proces ładowania danych uznaje się za udany.

Apache Airflow: praktyczne zastosowanie w automatyzacji ETL

Apache Airflow jest szeroko stosowany do organizowania złożonych procesów przetwarzania danych, umożliwiając budowanie elastycznych przenośników ETL. Apache Airflow może być uruchamiany za pomocą interfejsu WWW lub programowo za pomocą kodu Python (rys. 7.4-2). W interfejsie internetowym (rys. 7.4-3) administratorzy i programiści mogą wizualnie monitorować DAG, uruchamiać zadania i analizować wyniki wykonania.

Korzystając z DAG, można ustawić wyraźną sekwencję zadań, zarządzać zależnościami między nimi i automatycznie reagować na zmiany w danych źródłowych. Rozważmy przykład wykorzystania Airflow do automatyzacji przetwarzania raportów (rys. 7.4-2).



Rys. 7.4-2 ETL -koncepcja przenośnika do przetwarzania danych przy użyciu Apache Airflow.

Ten przykład (Rys. 7.4-2) uwzględnia DAG, który wykonuje kluczowe zadania w ramach ETL -conveyor:

■ **Read Excel -files (Extract):**

- Sekwencyjne przeglądanie wszystkich plików w danym katalogu.
- Odczyt danych z każdego pliku przy użyciu biblioteki pandas.
- Scala wszystkie dane w jedną ramkę DataFrame.

■ **Utwórz dokument PDF (Przekształć):**

- Przekształć scaloną DataFrame w HTML-table.
- Zapisz tabelę jako PDF (w wersji demonstracyjnej - przez HTML).

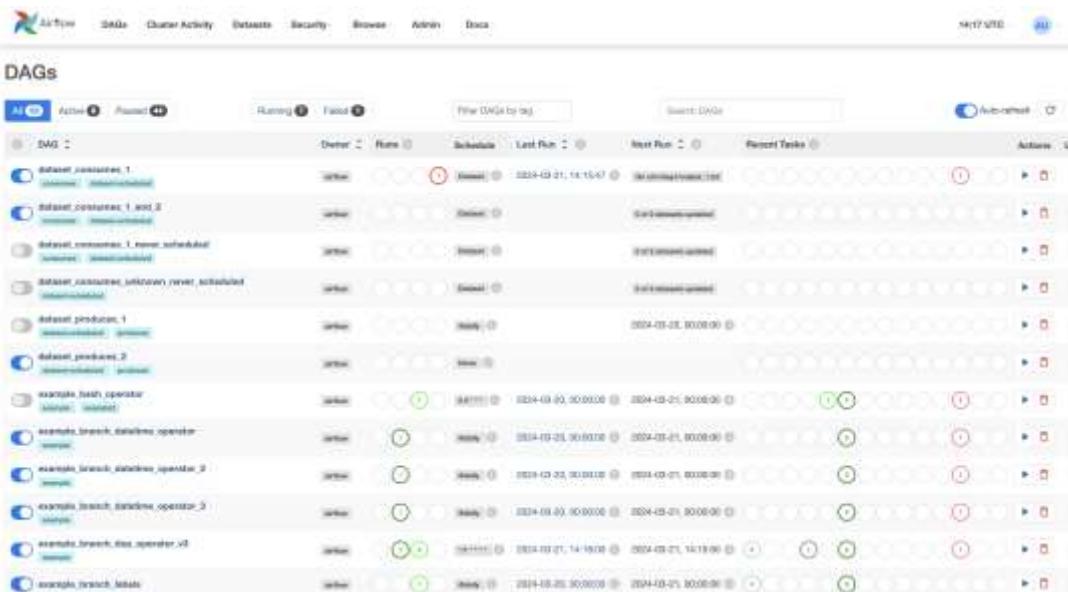
■ **Wysyłanie raportu e-mailem (Załaduj):**

- Apply EmailOperator to send PDF -document by email.

■ **Konfigurowanie DAG:**

- Definiowanie sekwencji zadań: wyodrębnianie danych → generowanie raportu → wysyłanie.
- Przypisanie harmonogramu uruchamiania (@monthly - pierwszy dzień każdego miesiąca).

Zautomatyzowany przykład ETL (Rys. 7.4-2) pokazuje, jak zbierać dane z plików Excel, tworzyć dokument PDF i wysyłać go pocztą elektroniczną. Jest to tylko jeden z wielu możliwych scenariuszy wykorzystania Airflow. Ten przykład można dostosować do dowolnego konkretnego zadania, aby uprościć i zautomatyzować przetwarzanie danych.



Rys. 7.4-3 Przegląd wszystkich DAG w środowisku z informacjami o ostatnich uruchomieniach.

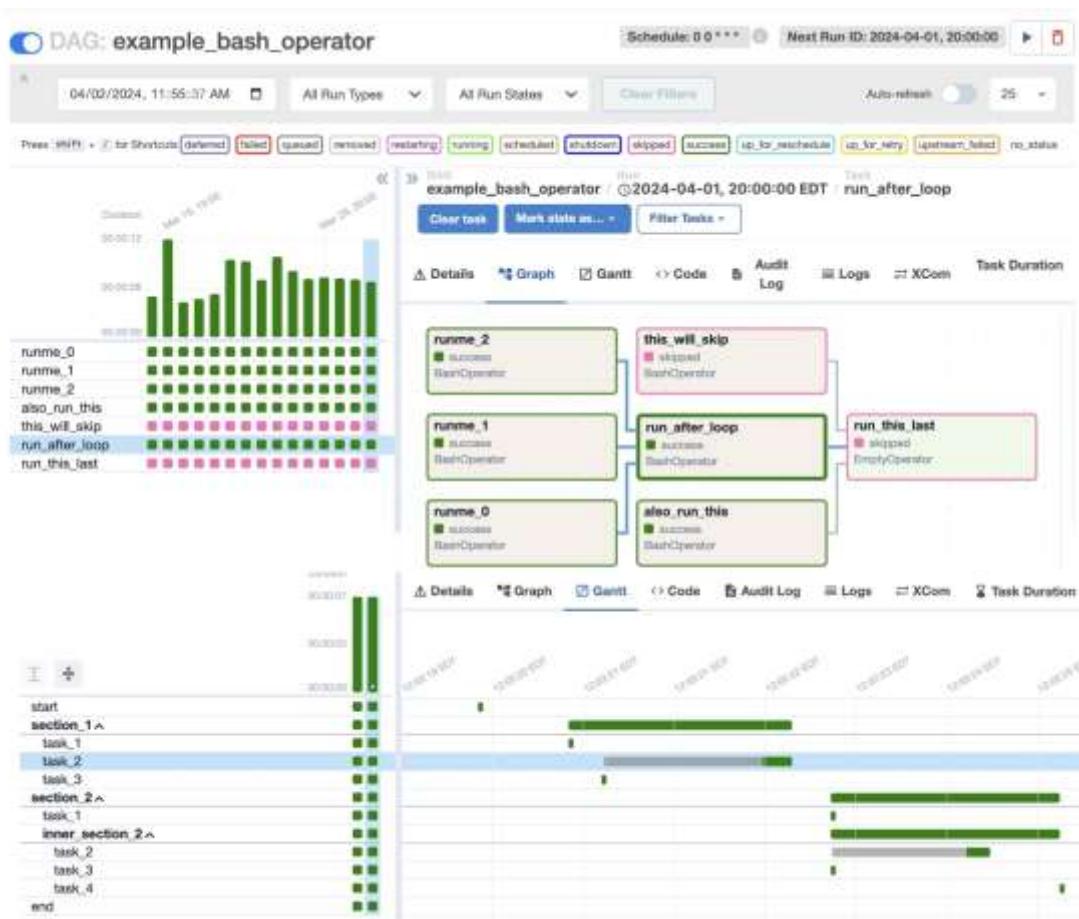
Interfejs internetowy Apache Airflow (Rys. 7.4-3) zapewnia kompleksowe środowisko wizualne do zarządzania przepływami danych. Wyświetla on DAG jako interaktywne grafy, w których węzły reprezentują zadania, a krawędzie reprezentują zależności między nimi, ułatwiając śledzenie złożonych przepływów danych. Interfejs zawiera pulpit nawigacyjny z informacjami na temat stanu wykonania zadania, historii uruchomień, szczegółowych dzienników i wskaźników wydajności. Administratorzy mogą ręcznie uruchamiać zadania, ponownie uruchamiać nieudane operacje, zawieszać DAG i konfigurować zmienne środowiskowe, a wszystko to za pośrednictwem intuicyjnego interfejsu użytkownika.

Taką architekturę można uzupełnić o walidację danych, powiadomienia o statusie wykonania, integrację z

zewnętrznymi API lub bazami danych. Airflow pozwala elastycznie dostosowywać DAG: dodawać nowe zadania, zmieniać ich kolejność, łączyć łańcuchy - co czyni go skutecznym narzędziem do automatyzacji złożonych procesów przetwarzania danych. Podczas uruchamiania DAG w interfejsie internetowym Airflow (rys. 7.4-3, rys. 7.4-4) można monitorować status zadań. System wykorzystuje oznaczenia kolorystyczne:

- Zielony - zadanie zostało pomyślnie ukończone.
- Żółty - proces jest w toku.
- Czerwony - błąd podczas wykonywania zadania.

W przypadku awarii (np. brakującego pliku lub uszkodzonej struktury danych) system automatycznie inicjuje wysyłanie powiadomienia.



Rys. 7.4-4 Apache Airflow znacznie upraszcza diagnostykę problemów, optymalizację procesów i współpracę zespołową nad złożonymi potokami przetwarzania danych.

Apache Airflow jest wygodny, ponieważ automatyzuje rutynowe zadania, eliminując potrzebę ich ręcznego wykonywania. Zapewnia niezawodność poprzez monitorowanie wykonywania procesów i natychmiastowe powiadamianie o błędach. Elastyczność systemu ułatwia dodawanie nowych zadań lub modyfikowanie istniejących, dostosowując przepływy pracy do zmieniających się wymagań.

Oprócz Apache Airflow istnieją podobne narzędzia do orkiestracji przepływów pracy. Na przykład darmowy Prefect o otwartym kodzie źródłowym (Rys. 7.3-5) oferuje prostszą składnię i lepiej integruje się z Pythonem, Luigi, opracowany przez Spotify, zapewnia podobną funkcjonalność i dobrze współpracuje z dużymi zbiorami danych. Warto również zwrócić uwagę na Kronos i Dagster, które oferują nowoczesne podejście do tworzenia Pipeline z naciskiem na modułowość i skalowalność. Wybór narzędzia do orkiestracji zadań zależy od konkretnych potrzeb projektu, ale wszystkie one pomagają zautomatyzować złożone procesy ETL danych.

Na szczególną uwagę zasługuje Apache NiFi, platforma open source, przeznaczona do strumieniowego przesyłania i routingu danych. W przeciwieństwie do Airflow, który koncentruje się na przetwarzaniu wsadowym i zarządzaniu zależnościami, NiFi koncentruje się na transformacji danych w czasie rzeczywistym i elastycznym routingu między systemami.

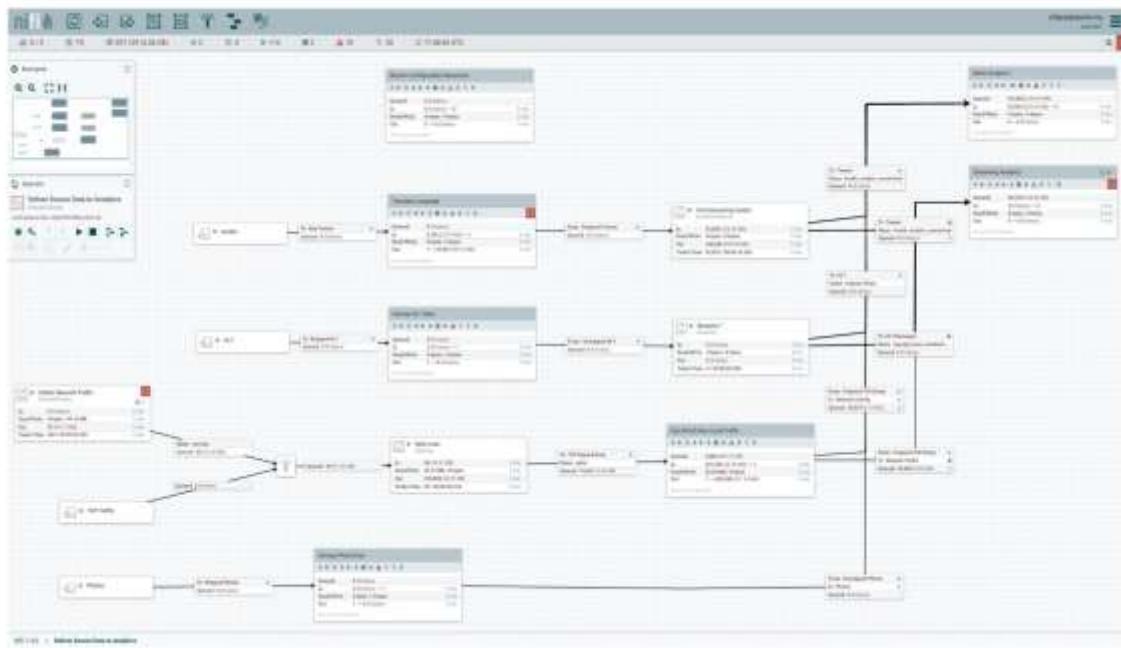
Apache NiFi do routingu i konwersji danych

Apache NiFi to potężna platforma open source, zaprojektowana do automatyzacji przepływu danych między różnymi systemami. Pierwotnie opracowana w 2006 roku przez amerykańską Agencję Bezpieczeństwa Narodowego (NSA) pod nazwą "Niagara Files" do użytku wewnętrznego. W 2014 r. projekt został udostępniony na zasadach open source i przeniesiony do Apache Software Foundation, stając się częścią ich inicjatyw transferu technologii [141].

Apache NiFi został zaprojektowany do zbierania, przetwarzania i przesyłania danych w czasie rzeczywistym. W przeciwieństwie do Airflow, który działa z zadaniami wsadowymi i wymaga dobrze zdefiniowanych harmonogramów, NiFi działa w trybie przetwarzania strumieniowego, umożliwiając ciągłe przesyłanie danych między różnymi usługami.

Apache NiFi jest idealny do integracji z urządzeniami IoT, czujnikami budowlanymi, systemami monitorowania i np. walidacji strumieniowej formatów CAD na serwerze, gdzie może być wymagana natychmiastowa reakcja na zmiany danych.

Dzięki wbudowanym narzędziom do filtrowania, transformacji i routingu, NiFi umożliwia standaryzację danych (Transform) przed ich przesaniem (Load) do systemów pamięci masowej lub systemów analitycznych. Jedną z jego głównych zalet jest wbudowane wsparcie bezpieczeństwa i kontrola dostępu, co czyni go niezawodnym rozwiązaniem do obsługi poufnych informacji.



Rys. 7.4-5 Graficzna reprezentacja przepływu danych w interfejsie Apache NiFi.

Apache NiFi wydajnie obsługuje strumieniowe przesyłanie danych w czasie rzeczywistym, filtrowanie i zadania routingu. Idealnie nadaje się do intensywnych technicznie scenariuszy, w których ważny jest stabilny transfer informacji między systemami i wysoka przepustowość.

Jednak gdy głównym celem jest integracja różnych usług, automatyzacja rutynowych operacji i szybkie konfigurowanie przepływów pracy bez głębokiej wiedzy programistycznej, pożąданie są rozwiązania o niskim progu wejścia i maksymalnej elastyczności. Jednym z takich narzędzi jest n8n - platforma klasy Low-Code /No-Code skoncentrowana na automatyzacji biznesu i wizualnej orkiestracji procesów.

n8n Low-Code, No-Code orkiestracja procesów

n8n to platforma Open Source Low-Code / No-Code do tworzenia zautomatyzowanych przepływów pracy, charakteryzująca się łatwością obsługi, elastycznością i możliwością szybkiej integracji z szeroką gamą usług zewnętrznych.

No-Code to metoda tworzenia produktów cyfrowych bez pisania kodu. Wszystkie elementy procesu - od logiki po interfejs - są realizowane wyłącznie za pomocą narzędzi wizualnych. Platformy No-Code są skierowane do użytkowników bez zaplecza technicznego i umożliwiają szybkie tworzenie automatyzacji, formularzy, integracji i aplikacji internetowych. Przykład: użytkownik konfiguruje automatyczne wysyłanie powiadomień lub integrację z Arkuszami Google za pomocą interfejsu "przeciągnij i upuść" bez wiedzy programistycznej.

Dzięki otwartemu oprogramowaniu i możliwościami lokalnego wdrażania, n8n w procesach automatyzacji i

tworzenia potoków ETL daje firmom pełną kontrolę nad ich danymi, zapewniając jednocześnie bezpieczeństwo i niezależność od dostawców usług w chmurze.

W przeciwnieństwie do Apache Airflow, który jest zorientowany na zadania obliczeniowe ze sztywną orkiestracją i wymaga znajomości języka Python, n8n zapewnia wizualny edytor, który umożliwia tworzenie skryptów bez konieczności znajomości języków programowania (Rys. 7.4-6). Choć jego interfejs umożliwia tworzenie zautomatyzowanych procesów bez pisania kodu (No-Code), w bardziej złożonych scenariuszach użytkownicy mogą dodawać własne funkcje JavaScript i Python w celu rozszerzenia możliwości (Low-Code).

Low-Code to podejście do tworzenia oprogramowania, w którym podstawowa logika aplikacji lub procesu jest tworzona przy użyciu interfejsu graficznego i elementów wizualnych, a kod programu jest używany tylko do dostosowywania lub rozszerzania funkcjonalności. Platformy Low-Code pozwalają znacznie przyspieszyć rozwój rozwiązań, angażując nie tylko programistów, ale także użytkowników biznesowych o podstawowych umiejętnościach technicznych.

Przykład: użytkownik może zbudować proces biznesowy z gotowych bloków i w razie potrzeby dodać własny skrypt w JavaScript lub Python.

Choć n8n jest pozycjonowany jako platforma o niskim progu wejścia, wymagana jest podstawowa wiedza programistyczna, zrozumienie technologii internetowych i umiejętności pracy z API. Elastyczność systemu pozwala na dostosowanie go do szerokiego zakresu zadań - od automatycznego przetwarzania danych po integrację z komunikatorami, urządzeniami IoT i usługami w chmurze.

Kluczowe funkcje i zalety korzystania z n8n:

- Opcje **open source** i lokalnego wdrożenia zapewniają pełną kontrolę danych, zgodność z przepisami bezpieczeństwa i niezależność od dostawców usług w chmurze.
- **Integracja z ponad 330 usługami**, w tym CRM, ERP, e-commerce, platformami w chmurze, komunikatorami i bazami danych.
- **Elastyczność scenariuszy**: od prostych powiadomień po złożone łańcuchy z przetwarzaniem API zadań, logiką decyzyjną i połączeniem usług AI.
- **Wsparcie dla JavaScript i Python**: użytkownicy mogą osadzać niestandardowy kod w razie potrzeby, rozszerzając możliwości automatyzacji.
- **Intuicyjny interfejs wizualny**: umożliwia szybką konfigurację i wizualizację wszystkich etapów procesu.

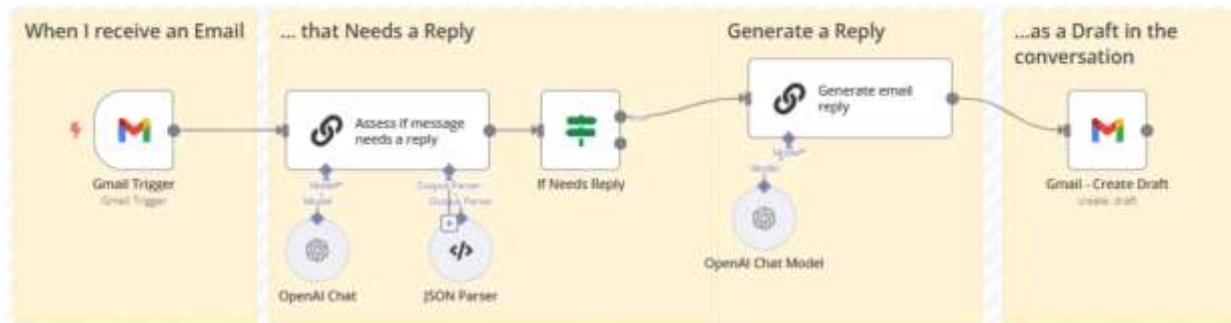
Platformy klasy Low-Code zapewniają narzędzia do tworzenia rozwiązań cyfrowych z minimalną ilością kodu, dzięki czemu są idealne dla zespołów, które nie mają głębokiej wiedzy technicznej, ale muszą zautomatyzować procesy.

W branży budowlanej n8n może być wykorzystywany do automatyzacji różnych procesów, takich jak integracja z systemami zarządzania projektami, sprawdzanie strumieni, pisanie gotowych raportów i listów, automatyczna aktualizacja danych dotyczących zapasów materiałów, wysyłanie powiadomień o statusie zadań do zespołów i wiele innych. Dostosowany Pipeline w n8n może wielokrotnie zmniejszyć liczbę operacji wykonywanych ręcznie, zmniejszyć prawdopodobieństwo wystąpienia błędów i przyspieszyć podejmowanie decyzji dotyczących realizacji projektu.

Możesz wybierać spośród prawie dwóch tysięcy gotowych, bezpłatnych i otwartych źródeł n8n Pipeline, dostępnych pod adresem: n8n.io/workflows, aby zautomatyzować zarówno przepływy pracy w budownictwie, jak i zadania osobiste, ograniczając rutynowe operacje.

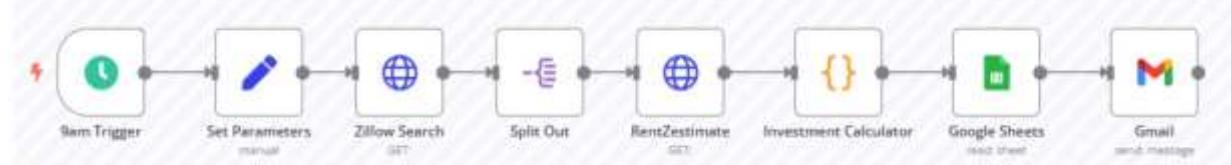
Weźmy na przykład jeden z gotowych szablonów Pipeline, dostępny za darmo na stronie n8n.io [142], który automatycznie tworzy wersje robocze odpowiedzi w Gmailu (rys. 7.4-6), pomagając użytkownikom, którzy otrzymują dużą liczbę wiadomości e-mail lub mają trudności z tworzeniem odpowiedzi.

Ten szablon n8n "Gmail AI Auto-Responder: Create draft responses to incoming emails" (Rys. 7.4-6) analizuje przychodzące wiadomości e-mail za pomocą LLM z ChatGPT, określa, czy potrzebna jest odpowiedź, generuje wersję roboczą z ChatGPT i konwertuje tekst na HTML i dodaje go do łańcucha wiadomości w Gmailu. Nie powoduje to automatycznego wysłania wiadomości e-mail, umożliwiającręczną edycję i zatwierdzenie odpowiedzi. Konfiguracja zajmuje około 10 minut i obejmuje konfigurację OAuth API Gmaila oraz integrację API OpenAI. Rezultatem jest wygodne i bezpłatne rozwiązanie do automatyzacji rutynowej komunikacji e-mail bez utraty kontroli nad treścią wiadomości.



Rys. 7.4-6 Zautomatyzowany proces generowania odpowiedzi e-mail przy użyciu n8n.

Innym przykładem automatyzacji za pomocą n8n jest znajdowanie świetnych ofert na rynku nieruchomości [143]. N8n Pipeline "Automate daily property deals using Zillow API, Google Sheets and Gmail", codziennie zbiera odpowiednie oferty pasujące do podanych kryteriów za pomocą Zillow API. Automatycznie oblicza kluczowe wskaźniki inwestycyjne (Cash on Cash ROI, Monthly Cash Flow, Down Payment), aktualizuje Google Sheets i wysyła raport podsumowujący na e-mail (Rys. 7.4-7), pozwalając inwestorom zaoszczędzić czas i szybko reagować na najlepsze oferty.



Rys. 7.4-7 Zautomatyzowany proces oceny atrakcyjności inwestycyjnej nieruchomości.

Jego elastyczność i rozszerzalność sprawiają, że n8n jest cennym narzędziem dla firm, które chcą dokonać cyfrowej transformacji i stać się bardziej konkurencyjne na rynku dzięki stosunkowo prostym i darmowym narzędziom open source.

Narzędzia takie jak Apache NiFi, Airflow i n8n można traktować jako trzy warstwy przetwarzania danych

(Rys. 7.4-8). NiFi zarządza przepływem danych, zapewniając ich dostarczanie i przekształcanie, Airflow orkiestruje wykonywanie zadań poprzez agregowanie danych w potoki przetwarzania, a n8n automatyzuje integrację z usługami zewnętrznymi i zarządza logiką biznesową.

	The main task	Approach
Apache NiFi	Streaming and data transformation	Real-time stream processing
Apache Airflow	Task orchestration, ETL pipelines	Batch planning, DAG processes
n8n	Integration, automation of business logic	Low-code visual orchestration

Rys. 7.4-8 Apache Airflow, Apache NiFi i n8n mogą być postrzegane jako trzy uzupełniające się warstwy nowoczesnej architektury zarządzania danymi.

Razem te darmowe i otwarte narzędzia potencjalnie tworzą przykład skutecznego ekosystemu zarządzania danymi i procesami w branży budowlanej, umożliwiając firmom wykorzystanie informacji do podejmowania decyzji i automatyzacji procesów.

Kolejne kroki: przejście od operacji ręcznych do rozwiązań opartych na analityce

Dzisiejsze firmy budowlane działają w środowisku dużej niepewności: zmieniające się ceny materiałów, opóźnione dostawy, niedobory siły roboczej i napięte terminy realizacji projektów. Korzystanie z pulpitów analitycznych, przenośników ETL i systemów BI pomaga firmom szybko identyfikować obszary problemowe, oceniać efektywność zasobów i przewidywać zmiany, zanim doprowadzą one do strat finansowych.

Podsumowując tę część, warto podkreślić główne praktyczne kroki, które pomogą zastosować omawiane technologie w codziennych zadaniach:

- Wdrażanie wizualizacji danych i pulpitów analitycznych
 - Opanowanie procesu tworzenia pulpitów nawigacyjnych w celu monitorowania kluczowych wskaźników wydajności (KPI).
 - Korzystaj z narzędzi do wizualizacji danych (Power BI, Tableau, Matplotlib, Plotly).
- Automatyzacja przetwarzania danych poprzez procesy ETL
 - Konfiguracja automatycznego gromadzenia danych z różnych źródeł (dokumentacja, tabele, CAD) za pomocą procesów ETL.
 - Organizowanie transformacji danych (np. sprawdzanie lub obliczanie wyrażeń regularnych) przy użyciu skryptów Python.

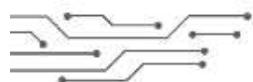
- Spróbuj skonfigurować automatyczne raportowanie PDF (lub DOC) za pomocą biblioteki FPDF, używając danych z plików Excel lub wyodrębniając informacje z innych dokumentów PDF

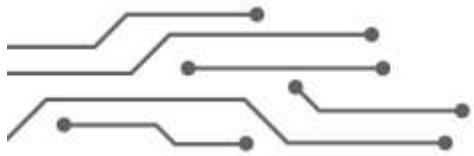
■ Wykorzystanie modeli językowych (LLM) do automatyzacji

- Korzystanie z dużych modeli językowych (LLM) w celu generowania kodu pomagającego wyodrębniać i analizować dane z nieustrukturyzowanych dokumentów.
- Zapoznaj się z narzędziem do automatyzacji n8n i zbadaj gotowe szablony i studia przypadków na ich stronie internetowej. Zidentyfikuj, które procesy z Twojej pracy mogą być w pełni zautomatyzowane przy użyciu podejścia No-Code/Low-Code.

Analityczne podejście do automatyzacji danych i procesów nie tylko skracą czas poświęcany na rutynowe operacje, ale także poprawia jakość podejmowania decyzji. Firmy, które wdrażają narzędzia do analizy wizualnej i ETL -conveyors zyskują możliwość szybkiego reagowania na zmiany

Automatyzacja procesów biznesowych za pomocą narzędzi takich jak n8n, Airflow i NiFi to tylko pierwszy krok w kierunku cyfrowej dojrzałości. Kolejnym krokiem jest wysokiej jakości przechowywanie i zarządzanie danymi, które stanowią podstawę automatyzacji. W części 8 przyjrzymy się dogłębnie, w jaki sposób firmy budowlane mogą zbudować zrównoważoną architekturę przechowywania danych, przechodząc od chaosu dokumentów i plików wieloformatowych do scentralizowanej pamięci masowej i platform analitycznych.





VII CZĘŚĆ

PRZECHOWYWANIE I ZARZĄDZANIE DANYMI W BUDOWNICTWIE

W części 8 omówiono nowoczesne technologie przechowywania i zarządzania danymi dla branży budowlanej. Analizuje wydajne formaty do obsługi dużych ilości informacji - od prostych CSV i XLSX do bardziej wydajnych Apache Parquet i ORC ze szczególnym porównaniem ich możliwości i ograniczeń. Omówiono koncepcje hurtowni danych (DWH), jezior danych) i ich rozwiązań hybrydowych (Data Lakehouse), a także zasady zarządzania danymi) i minimalizmu danych). Szczegółowo omówiono problemy związane z bagnem danych (Data Swamp) oraz strategie zapobiegania chaosowi w systemach informatycznych. Przedstawiono nowe podejścia do pracy z danymi, w tym wektorowe bazy danych i ich zastosowanie w konstrukcji poprzez koncepcję Bounding Box. W tej części omówiono również metodologie DataOps i VectorOps jako nowe standardy organizacji przepływu pracy z danymi.

ROZDZIAŁ 7.1.

INFRASTRUKTURA DANYCH: OD FORMATÓW PRZECHOWYWANIA DO REPOZYTORIÓW

Atomy danych: podstawa skutecznego zarządzania informacjami

Wszystko we Wszechświecie składa się z najmniejszych bloków budulcowych - atomów i cząsteczek, a z czasem wszystkie żywe i nieożywione rzeczy nieuchronnie powracają do tego stanu początkowego. W naturze proces ten zachodzi z zadziwiającą prędkością, którą staramy się przenieść na procesy kontrolowane przez człowieka.

W lesie wszelkie żywne organizmy są ostatecznie przekształcane w pożywną substancję, która służy jako podstawa dla nowych roślin. Rośliny te z kolei stają się pożywieniem dla nowych żywych istot zbudowanych z tych samych atomów, które stworzyły Wszechświat miliony lat temu.

W świecie biznesu ważne jest również rozbicie złożonych, wielowarstwowych struktur na ich najbardziej fundamentalne, minimalnie przetworzone jednostki - podobnie jak atomy i cząsteczki w przyrodzie. Pozwala to na efektywne przechowywanie i zarządzanie atomami danych, przekształcając je w bogaty, żywy fundament, który staje się kluczowym zasobem dla rozwoju analityki i jakości decyzji.



Rys. 8.1-1 Analiza i podejmowanie decyzji opiera się na ponownie wykorzystanych danych, które zostały kiedyś przetworzone i zapisane.

Kompozycje muzyczne składają się z nut, które łączą się, tworząc złożone utwory muzyczne, podczas gdy

słowa są tworzone z prymitywnej jednostki, litery-dźwięku. Niezależnie od tego, czy jest to natura, nauka, ekonomia, sztuka czy technologia, świat wykazuje niezwykłą jedność i harmonię w dążeniu do destrukcji, struktury, jazdy na rowerze i tworzenia. Podobnie, procesy w systemach kalkulacji kosztów są podzielone na małe ustrukturyzowane jednostki - pozycje zasobów - na poziomie kalkulacji kosztów i harmonogramów. Jednostki te, podobnie jak notatki, są następnie wykorzystywane do tworzenia bardziej złożonych obliczeń i harmonogramów. Ta sama zasada jest wykorzystywana przez systemy projektowania wspomaganego komputerowo, w których złożone projekty architektoniczne i inżynierijne są budowane z podstawowych elementów - pojedynczych elementów i komponentów bibliotecznych, z których tworzony jest kompletny model 3D projektu złożonego budynku lub konstrukcji.

Koncepcja cykliczności i struktury właściwa naturze i nauce znajduje również odzwierciedlenie we współczesnym świecie danych. Tak jak w naturze wszystkie żywne istoty powracają do atomów i cząsteczek, tak w świecie nowoczesnych narzędzi do przetwarzania danych, informacje mają tendencję do powracania do swojej najbardziej prymitywnej formy.

Najmniejsze elementy, z ich skończoną niepodzielnością, są podstawowymi elementami składowymi procesów biznesowych. Ważne jest, aby od samego początku dokładnie rozważyć, w jaki sposób gromadzić, strukturyzować (rozkładać na atomy) i przechowywać te małe bloki konstrukcyjne z różnych źródeł. Organizowanie i przechowywanie danych to nie tylko kwestia rozbicia ich na części składowe. Równie ważne jest zapewnienie, że są one zintegrowane i przechowywane w ustrukturyzowany sposób, dzięki czemu dane mogą być łatwo wyszukiwane, analizowane i wykorzystywane do podejmowania decyzji, gdy tylko są potrzebne.

Aby wydajnie przetwarzać informacje, format i metody przechowywania danych muszą być starannie dobrane - podobnie jak gleba musi być przygotowana do wzrostu drzew. Magazyny danych muszą być zorganizowane tak, aby zapewnić wysoką jakość i trafność informacji, eliminując zbędne lub nieistotne dane. Im lepiej zorganizowana jest ta "gleba informacyjna", tym szybciej i dokładniej użytkownicy mogą znaleźć właściwe dane i rozwiązać problemy analityczne.

Przechowywanie informacji: pliki lub dane

Hurtownie danych pozwalają firmom gromadzić i łączyć informacje z różnych systemów, tworząc jedno centrum do dalszej analizy. Zebrane dane historyczne umożliwiają nie tylko głębszą analizę procesów, ale także identyfikację wzorców, które mogą wpływać na wyniki biznesowe.

Założymy, że firma pracuje nad kilkoma projektami jednocześnie. Inżynier chce wiedzieć, ile betonu już wyłano, a ile jeszcze pozostało do kupienia. W tradycyjnym podejściu musiałby ręcznie przeszukać serwer i otworzyć kilka tabel szacunkowych, porównać je z certyfikatami ukończonych prac i sprawdzić aktualne stany magazynowe. Zajmuje to godziny, a nawet dni. Nawet w przypadku procesów ETL i automatycznych skryptów, zadanie pozostaje częściowo manualne: inżynier nadal musi ręcznie określić ścieżkę do folderów lub określonych plików na serwerze. Zmniejsza to ogólny efekt automatyzacji, ponieważ nadal zajmuje cenny czas pracy.

Przechodząc na zarządzanie danymi, zamiast pracować z systemem plików serwera, inżynier uzyskuje dostęp do ujednoliconej struktury przechowywania, w której informacje są aktualizowane w czasie rzeczywistym. Pojedyncze zapytanie - w formie kodu, zapytania SQL lub nawet wywołania agenta LLM -

może natychmiast dostarczyć dokładnych danych na temat bieżących sald, wykonanej pracy i nadchodzących dostaw, jeśli dane zostały wcześniej przygotowane i połączone w hurtowni danych, gdzie nie ma potrzeby wędrowania po folderach, otwierania dziesiątek plików i ręcznego porównywania wartości.

Przez długi czas firmy budowlane korzystały z dokumentów PDF, rysunków DWG, modeli RVT oraz setek i tysięcy tabel Excel i innych różnych formatów, które są przechowywane w określonych folderach na serwerach firmy, co utrudnia wyszukiwanie informacji, ich sprawdzanie i analizowanie. W rezultacie pliki pozostałe po zakończeniu projektów są najczęściej przenoszone z powrotem na serwer do folderów archiwalnych, które praktycznie nie są wykorzystywane w przyszłości. Takie tradycyjne przechowywanie danych oparte na plikach traci na znaczeniu wraz ze wzrostem przepływu danych, ze względu na podatność na błędy ludzkie.

Plik to po prostu izolowany kontener, w którym przechowywane są dane. Pliki są tworzone dla ludzi, a nie dla systemów, więc wymagają ręcznego otwierania, odczytu i interpretacji.

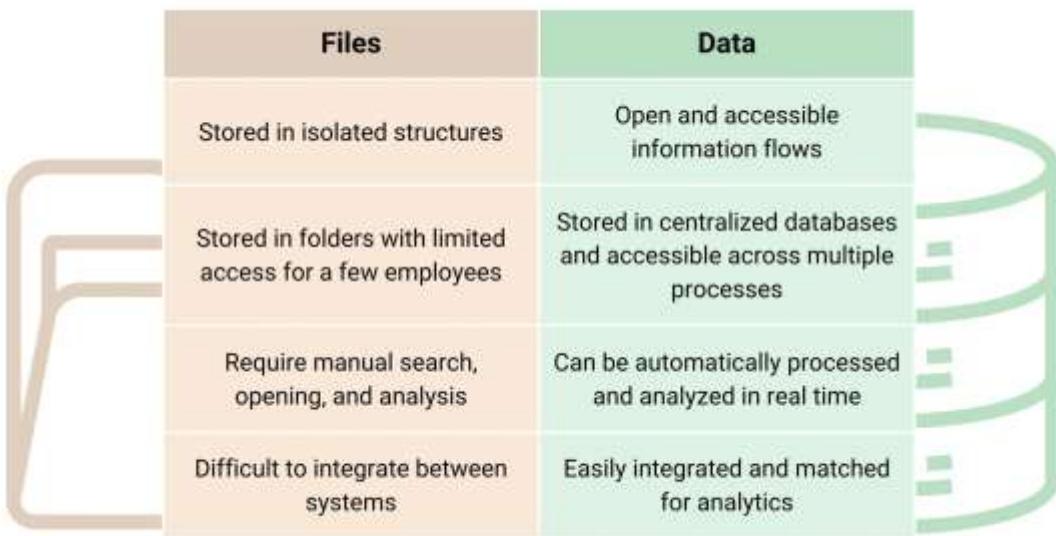
Przykładem może być tabela Excel, dokument PDF lub Rys. CAD, który należy otworzyć w określonym narzędziu, aby uzyskać dostęp do żądanych informacji. Bez ustrukturyzowanego wyszukiwania i przetwarzania, zawarte w nich informacje pozostają niewykorzystane.

Z kolei **dane** to informacje nadające się do odczytu maszynowego, które są automatycznie łączone, aktualizowane i analizowane. W pojedynczej hurtowni danych (np. bazie danych, DWH lub Data Lake) informacje są reprezentowane w postaci tabel, rekordów i relacji. Umożliwia to jednolite przechowywanie, zautomatyzowane zapytania, analizy wartości i raportowanie w czasie rzeczywistym.

Używanie danych zamiast plików (rys. 8.1-1) eliminuje ręczny proces wyszukiwania i ujednolicza przetwarzanie. Firmy, które już wdrożyły takie podejście, zyskują przewagę konkurencyjną dzięki szybkości dostępu do informacji i możliwości szybkiego zintegrowania ich z procesami biznesowymi.

Przejście od korzystania z plików do danych jest nieuniknioną zmianą, która ukształtuje przyszłość branży budowlanej.

Każda firma z branży budowlanej stanie przed kluczowym wyborem: nadal przechowywać informacje w różnych plikach i silosach, które muszą być odczytywane przez ludzi za pomocą specjalnych programów lub przekształcać je na pierwszych etapach przetwarzania w ustrukturyzowane dane, tworząc jedną zintegrowaną cyfrową podstawę do zautomatyzowanego zarządzania projektami.



Rys. 8.1-1 Ewolucja przepływu informacji: od izolowanych plików do zintegrowanych danych.

Wraz z eksplozją informacji, tradycyjne metody przechowywania i przetwarzania plików stają się coraz mniej wydajne. W branży budowlanej, podobnie jak w innych sektorach, nie wystarczy już polegać na rozproszonych folderach z różnymi formatami plików lub niepołączonych bazach danych.

Firmy, które chcą pozostać konkurencyjne w erze cyfrowej, będą nieuchronnie przechodzić na zintegrowane platformy cyfrowe, wykorzystywać technologie Big Data i zautomatyzowane systemy analityczne.

Przejście od przechowywania opartego na plikach do przepływów pracy opartych na danych będzie wymagało ponownego przemyślenia podejścia do zarządzania informacjami i świadomego wyboru formatów odpowiednich do dalszej integracji w skoncentrowanych repozytoriach. Wybór ten określi, jak wydajnie można przetwarzać dane, jak szybko można uzyskać do nich dostęp i jak łatwo można je zintegrować z cyfrowymi procesami firmy.

Przechowywanie dużych ilości danych: analiza popularnych formatów i ich skuteczności

Formaty przechowywania danych odgrywają kluczową rolę w skalowalności, niezawodności i wydajności infrastruktury analitycznej. Do analizy i przetwarzania danych - takich jak filtrowanie, grupowanie i agregacja - w naszych przykładach wykorzystano Pandas DataFrame - popularną strukturę do pracy z danymi w pamięci RAM.

Pandas DataFrame nie ma jednak własnego formatu przechowywania danych, więc po zakończeniu przetwarzania dane są eksportowane do jednego z zewnętrznych formatów - najczęściej CSV lub XLSX. Te formaty tabelaryczne są łatwe do wymiany i kompatybilne z większością systemów zewnętrznych, ale mają szereg ograniczeń: niską wydajność przechowywania, brak kompresji i słabą obsługę wersjonowania:

- **CSV** (Comma-Separated Values): prosty format tekstowy szeroko obsługiwany przez różne platformy i narzędzia. Jest łatwy w użyciu, ale nie obsługuje złożonych typów danych i kompresji.

- **XLSX** (Excel Open XML Spreadsheet): format pliku programu Microsoft Excel, który obsługuje zaawansowane funkcje, takie jak formuły, wykresy i stylizacja. Chociaż jest przydatny do ręcznej analizy danych i wizualizacji, nie jest zoptymalizowany do przetwarzania danych na dużą skalę.

Oprócz popularnych tabelarycznych formatów XLSX i CSV, istnieje kilka popularnych formatów do efektywnego przechowywania danych strukturalnych (Rys. 8.1-2), z których każdy ma unikalne zalety w zależności od konkretnych wymagań dotyczących przechowywania i analizy danych:

- **Apache Parquet**: kolumnowy format przechowywania danych zoptymalizowany do użytku w systemach analizy danych. Oferuje wydajną kompresję danych i schematy kodowania, dzięki czemu idealnie nadaje się do złożonych struktur danych i przetwarzania dużych zbiorów danych.
- **Apache ORC (Optimised Row Columnar)**: Podobnie jak Parquet, ORC zapewnia wysoką kompresję i wydajne przechowywanie danych. Jest zoptymalizowany pod kątem ciężkich operacji odczytu i dobrze nadaje się do przechowywania jezior danych.
- **JSON (JavaScript Object Notation)**: chociaż JSON nie jest tak wydajny pod względem przechowywania danych w porównaniu do formatów binarnych, takich jak Parquet lub ORC, jest bardzo przystępny i łatwy w obsłudze, dzięki czemu idealnie nadaje się do skryptów, w których ważna jest czytelność i kompatybilność z siecią.
- **Feather**: szybki, lekki i łatwy w użyciu, zorientowany analitycznie, binarny kolumnowy format przechowywania danych. Został zaprojektowany do wydajnego przesyłania danych między Pythonem (Pandas) i R, co czyni go doskonałym wyborem dla projektów obejmujących te środowiska programistyczne.
- **HDF5 (Hierarchical Data Format version 5)**: przeznaczony do przechowywania i organizowania dużych ilości danych. Obsługuje szeroki zakres typów danych i dobrze nadaje się do pracy ze złożonymi zbiorami danych. HDF5 jest szczególnie popularny w obliczeniach naukowych ze względu na jego zdolność do wydajnego przechowywania i uzyskiwania dostępu do dużych zbiorów danych.



	XLSX	CSV	Apache Parquet	HDF5	Pandas DataFrame
Storage	Tabular	Tabular	Columnar	Hierarchical	Tabular
Usage	Office tasks, data presentation	Simple data exchange	Big data, analytics	Scientific data, large volumes	Data analysis, manipulation
Compression	Built-in	None	High	Built-in	None (in-memory)
Performance	Low	Medium	High	High	High (memory dependent)
Complexity	High (formatting, styles)	Low	Medium	Medium	Low
Data Type Support	Limited	Very limited	Extended	Extended	Extended
Scalability	Low	Low	High	High	Medium (memory limited)

Rys. 8.1-2 Porównanie formatów danych pokazujące główne różnice w aspektach przechowywania i przetwarzania.

Aby przeprowadzić analizę porównawczą formatów używanych na etapie ładowania procesu ETL, utworzono tabelę pokazującą rozmiary plików i czasy odczytu (rys. 8.1-3). W badaniu wykorzystano pliki z identycznymi danymi: tabela zawierała 10 000 wierszy i 10 kolumn wypełnionych losowymi wartościami.

W badaniu uwzględniono następujące formaty przechowywania danych: CSV, Parquet, XLSX i HDF5, a także ich skompresowane wersje w archiwach ZIP. Surowe dane zostały wygenerowane przy użyciu biblioteki NumPy i przedstawione jako struktura Pandas DataFrame. Proces testowania składał się z następujących kroków:

- Zapisywanie plików: ramka danych jest zapisywana w czterech różnych formatach: CSV, Parquet, XLSX i HDF5. Każdy format ma unikalne cechy w sposobie przechowywania danych, wpływając na rozmiar pliku i szybkość odczytu.
- Kompresja plików ZIP: aby przeanalizować skuteczność standardowej kompresji, każdy plik został dodatkowo skompresowany do archiwum ZIP.
- Odczyt pliku (ETL - Load): czas odczytu został zmierzony dla każdego pliku po rozpakowaniu go z ZIP. Pozwala to oszacować szybkość dostępu do danych po wyodrębnieniu z archiwum.

Należy zauważyć, że Pandas DataFrame nie był używany bezpośrednio w analizie rozmiaru lub czasu odczytu, ponieważ sam w sobie nie reprezentuje formatu przechowywania danych. Służył jedynie jako struktura pośrednia do generowania i późniejszego zapisywania danych w różnych formatach.



Rys. 8.1-3 Porównanie formatów pamięci masowej pod względem rozmiaru i szybkości odczytu.

Pliki CSV i HDF5 wykazują (Rys. 8.1-3) wysoką wydajność kompresji, znacznie zmniejszając ich rozmiar po spakowaniu w ZIP, co może być szczególnie przydatne w scenariuszach wymagających optymalizacji przechowywania. Z drugiej strony pliki XLSX są praktycznie nieskompresowalne, a ich rozmiar w ZIP pozostaje porównywalny z oryginałem, co czyni je mniej korzystnymi do wykorzystania w dużych ilościach danych lub w środowiskach, w których ważna jest szybkość dostępu do danych. Ponadto czas odczytu XLSX jest znacznie wyższy w porównaniu do innych formatów, co czyni go mniej korzystnym dla szybkich operacji odczytu danych. Apache Parquet wykazał wysoką wydajność w zadaniach analitycznych i dużych wolumenach danych ze względu na swoją kolumnową strukturę.

Optymalizacja pamięci masowej za pomocą Apache Parquet

Jednym z popularnych formatów przechowywania i przetwarzania dużych zbiorów danych jest Apache Parquet. Format ten został zaprojektowany specjalnie do przechowywania kolumnowego (podobnie jak Pandas), co pozwala znacznie zmniejszyć ilość zajmowanej pamięci i zwiększyć szybkość zapytań analitycznych. W przeciwieństwie do tradycyjnych formatów, takich jak CSV i XLSX, Parquet obsługuje natywną kompresję i jest zoptymalizowany pod kątem systemów Big Data, w tym Spark, Hadoop i przechowywania danych w chmurze.

Kluczowe funkcje Parquet obejmują obsługę kompresji i kodowania danych, co znacznie zmniejsza rozmiar pamięci masowej i przyspiesza operacje odczytu danych, działając bezpośrednio na żądanych kolumnach, a nie na wszystkich wierszach danych.

Aby zilustrować, jak łatwo jest uzyskać niezbędny kod do konwersji danych na Apache Parquet, użyjmy LLM.

- 👉 Wyślij prośbę tekstową na czat LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN):

Napisz kod, aby zapisać dane z Pandas DataFrame do Apache Parquet. ↵

- 👉 Odpowiedź LLM :

```

1 import pandas as pd
2
3 # Create a DataFrame with construction data
4 construction_data = {
5     'project_id': [1, 2, 3, 4],
6     'project_name': ['Tower A', 'Bridge B', 'Road C', 'Tunnel D'],
7     'location': ['City X', 'City Y', 'City Z', 'City W'],
8     'cost_million_usd': [200, 150, 100, 300]
9 }
10 df_projects = pd.DataFrame(construction_data)
11
12 # Write the DataFrame to the parquet file
13 parquet_path_projects = '/tmp/construction_data/projects.parquet'
14 df_projects.to_parquet(parquet_path_projects)
15
16 print("Construction project data saved in Lakehouse").
17

```

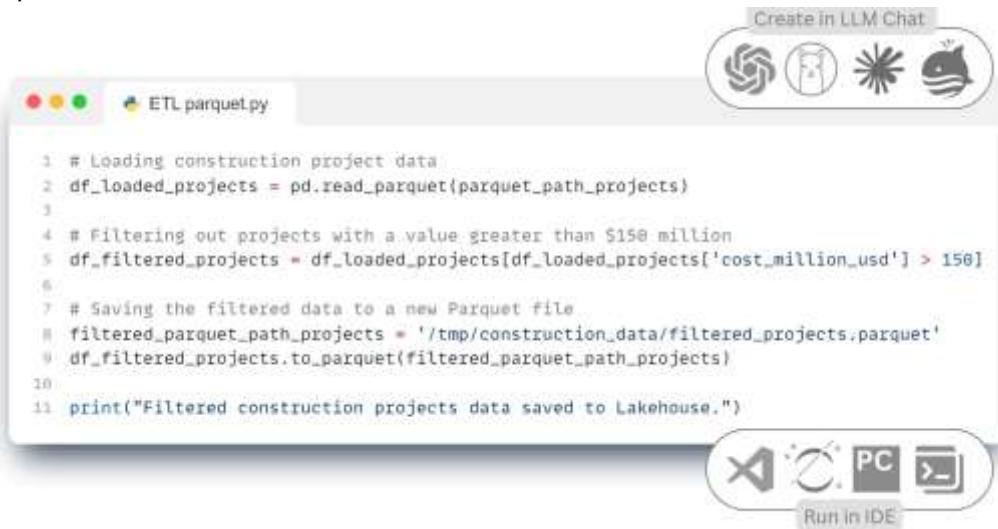
Rys. 8.1-4 Przenoszenie danych ramki danych z pamięci RAM do formatu wydajnego pod względem przechowywania Apache Parquet przy użyciu kilku wierszy Pythona .

Następny przykład: zasymulujmy proces ETL z danymi przechowywanymi w formacie Parquet w celu filtrowania projektów według określonej wartości jednego z atrybutów "cost_million_usd" (rys. 8.1-4).

- 👉 Gdy czat będzie kontynuowany, wyślij wiadomość tekstową na adres LLM:

Napisz kod, w którym chcemy filtrować dane w tabeli i zapisywać tylko te projekty (wiersze tabeli) z danych Apache Parquet, których koszt (parametr cost_million_usd) przekracza 150 milionów dolarów. ↴

💡 Odpowiedź LLM :



```

1 # Loading construction project data
2 df_loaded_projects = pd.read_parquet(parquet_path_projects)
3
4 # Filtering out projects with a value greater than $150 million
5 df_filtered_projects = df_loaded_projects[df_loaded_projects['cost_million_usd'] > 150]
6
7 # Saving the filtered data to a new Parquet file
8 filtered_parquet_path_projects = '/tmp/construction_data/filtered_projects.parquet'
9 df_filtered_projects.to_parquet(filtered_parquet_path_projects)
10
11 print("Filtered construction projects data saved to Lakehouse.")

```

The screenshot shows a Jupyter Notebook cell with the title "ETL parquet.py". The cell contains Python code for reading a Parquet file, filtering it to exclude projects costing more than \$150 million, and then saving the filtered data back to a new Parquet file. Below the code cell are icons for "Create in LLM Chat", "Run in IDE", and "Run in Jupyter".

Rys. 8.1-5 Proces ETL podczas pracy z danymi w formacie Apache Parquet wygląda tak samo, jak w przypadku innych formatów strukturalnych.

Korzystanie z formatu Parquet (w odniesieniu do XLSX, CSV itp.) znacznie zmniejsza ilość przechowywanych informacji i przyspiesza operacje wyszukiwania. Dzięki temu doskonale nadaje się zarówno do przechowywania, jak i analizowania danych. Parquet integruje się z różnymi systemami przetwarzania, zapewniając wydajny dostęp w architekturach hybrydowych.

Jednak wydajny format przechowywania to tylko jeden z elementów kompletnego środowiska danych. Aby stworzyć zrównoważone i skalowalne środowisko, wymagana jest dobrze zaprojektowana architektura zarządzania danymi. Systemy klasy DWH (Data Warehouse) spełniają tę funkcję. Zapewniają one agregację danych z heterogenicznych źródeł, przejrzystość procesów biznesowych oraz możliwość kompleksowej analizy z wykorzystaniem narzędzi BI i algorytmów uczenia maszynowego.

DWH: Hurtownia danych hurtownie danych

Podobnie jak format Parquet jest zoptymalizowany pod kątem wydajnego przechowywania dużych ilości informacji, Hurtownia Danych jest zoptymalizowana pod kątem integracji i strukturyzacji danych w celu wspierania analityki, prognozowania i podejmowania decyzji zarządczych.

W dzisiejszych firmach dane pochodzą z wielu różnych źródeł: systemów ERP, CAFM, CPM, CRM, księgowości i zarządzania magazynem, cyfrowych modeli CAD budynków, czujników IoT i innych rozwiązań. Aby uzyskać całościowy obraz, nie wystarczy po prostu gromadzić dane - należy je zorganizować, ustandaryzować i scentralizować w jednym repozytorium. To jest dokładnie to, co robi DWH

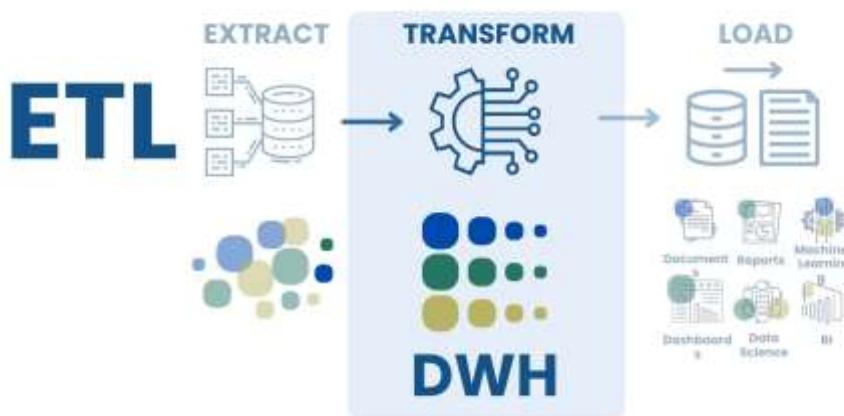
- scentralizowany system przechowywania, który pozwala agregować informacje z różnych źródeł, strukturyzować je i udostępniać do analizy i zarządzania strategicznego.

DWH (Data Warehouse) to scentralizowany system hurtowni danych, który agreguje informacje z wielu źródeł, strukturyzuje je i udostępnia do analizy i raportowania.

W wielu firmach dane są rozproszone w różnych systemach, które omówiliśmy w pierwszych częściach książki (Rys. 1.2-4). Hurtownia danych DWH integruje te źródła, zapewniając pełną przejrzystość i wiarygodność informacji. Hurtownia danych DWH to wyspecjalizowana baza danych (duża baza danych), która gromadzi, przetwarza i przechowuje dane z wielu źródeł. Główne cechy DWH to:

- **Korzystanie z procesów ETL** (Extract, Transform, Load) - wyodrębnianie danych ze źródeł, czyszczenie ich, przekształcanie, ładowanie do repozytorium i automatyzacja tych procesów, które zostały omówione w siódmej części książki.
- **Granularność danych** - dane w DWH mogą być przechowywane zarówno w formie zagregowanej (raporty zbiorcze), jak i granularnej (dane surowe). Począwszy od 2024 r. to dostawcy CAD-zaczęli mówić o danych granularnych [125], by móc wskazując, że branża przygotowuje się do przejścia na wyspecjalizowane przechowywanie w chmurze do obsługi cyfrowych danych modeli budynków.
- **Wspieranie analityki i predykcji** - hurtownie danych stanowią podstawę dla narzędzi BI, analizy Big Data i uczenia maszynowego.

DWH służy jako podstawa analityki biznesowej, umożliwiając analizę kluczowych wskaźników wydajności, prognozowanie sprzedaży, zakupów i kosztów oraz zautomatyzowane raportowanie i wizualizację danych (Rys. 8.1-6).



Rys. 8.1-6 W procesie ETL DWH może działać jako centralne repozytorium, w którym dane wyodrębnione z różnych systemów przechodzą etapy transformacji i wyładunku.

DWH odgrywa kluczową rolę w integracji, oczyszczaniu i strukturyzacji informacji, zapewniając solidne podstawy dla analityki biznesowej i procesów decyzyjnych. Jednak w dzisiejszym środowisku, w którym wolumeny danych szybko rosną, a źródła danych stają się coraz bardziej zróżnicowane, tradycyjne podejście DWH do przechowywania informacji często wymaga rozszerzenia w postaci ELT i Data Lake

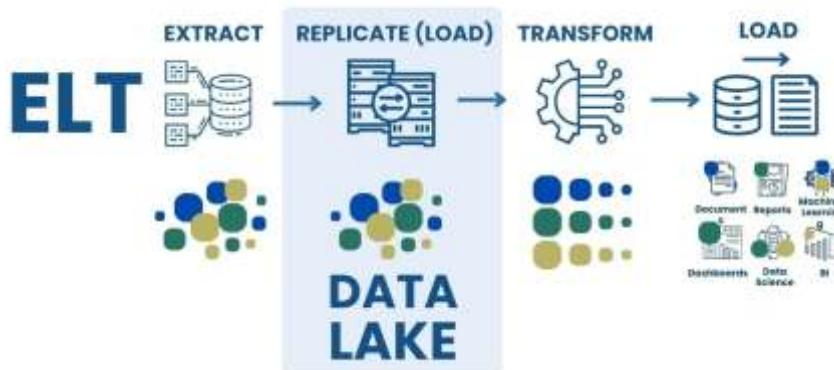
Data Lake - ewolucja ETL do ELT: od tradycyjnego czyszczenia do elastycznego przetwarzania

Klasyczne DWH - hurtownie danych, zaprojektowane do przechowywania ustrukturyzowanych danych w formacie zoptymalizowanym pod kątem zapytań analitycznych, napotkały ograniczenia w obsłudze danych nieustrukturyzowanych i skalowalności. W odpowiedzi na te wyzwania pojawiły się Data Lakes, oferujące elastyczne przechowywanie dużych ilości heterogenicznych danych.

Data Lake oferuje alternatywne podejście DWH, które umożliwia pracę z nieustrukturyzowanymi, częściowo ustrukturyzowanymi i nieprzetworzonymi danymi bez wcześniejszego sztywnego schematu. Ta metoda przechowywania danych jest często istotna dla przetwarzania danych w czasie rzeczywistym, uczenia maszynowego i zaawansowanej analityki. W przeciwieństwie do DWH, które strukturyzuje i agreguje dane przed załadowaniem, Data Lake pozwala na przechowywanie informacji w ich surowej postaci, zapewniając w ten sposób elastyczność i skalowalność

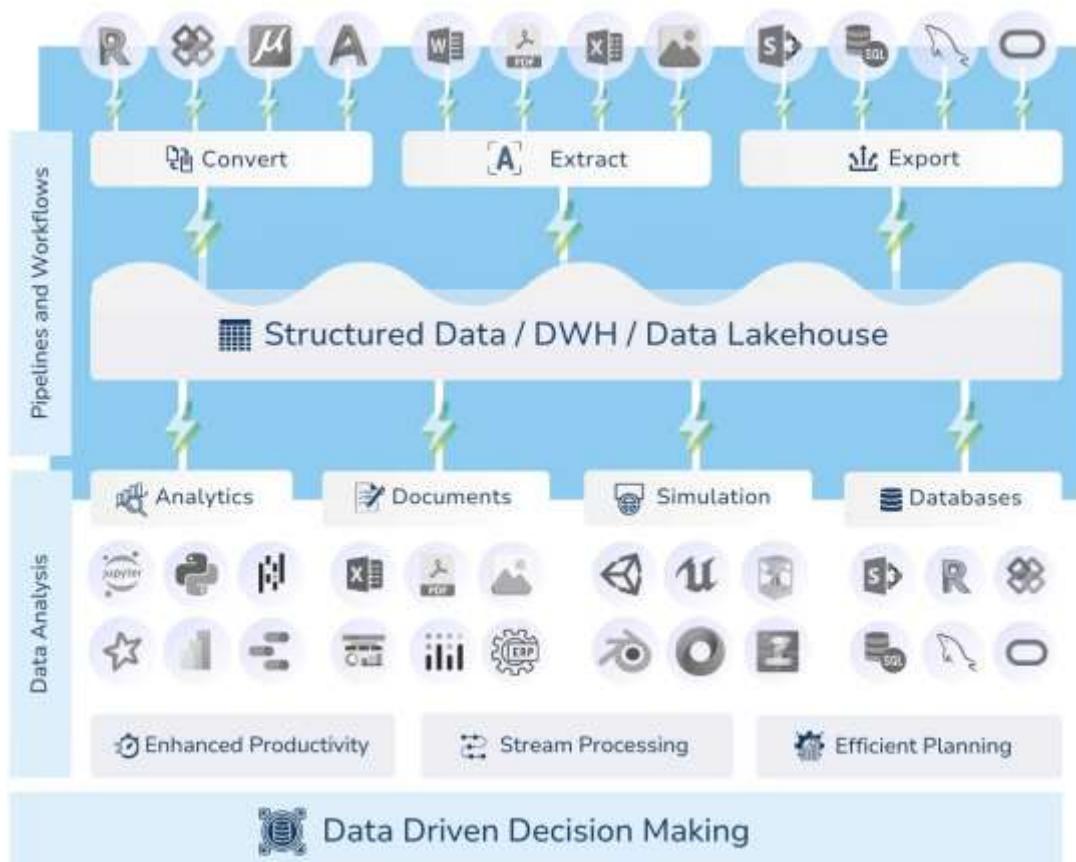
To właśnie frustracja tradycyjnymi hurtowniami danych (RDBMS, DWH) i zainteresowanie "big data" doprowadziły do powstania jezior danych, gdzie zamiast skomplikowanego ETL, dane są teraz po prostu ładowane do repozytorium o luźnej strukturze, a ich przetwarzanie odbywa się już na etapie analizy:

- W tradycyjnych hurtowniach danych dane są zazwyczaj wstępnie przetwarzane, przekształcane i czyszczone (ETL - Extract, Transform, Load) przed załadowaniem do hurtowni (Rys. 8.1-6). Oznacza to, że dane są ustrukturyzowane i zoptymalizowane pod kątem konkretnych przyszłych zadań analitycznych i raportowania. Nacisk kładziony jest na utrzymanie wysokiej wydajności zapytań i integralności danych. Podejście to może być jednak kosztowne i mniej elastyczne pod względem integracji nowych typów danych i szybko zmieniających się schematów danych.
- Z drugiej strony, jeziora danych są przeznaczone do przechowywania dużych ilości nieprzetworzonych danych w ich oryginalnym formacie (Rys. 8.1-7). Proces ETL (Extract, Transform, Load), jest zastępowany przez ELT (Extract, Load, Transform), w którym dane są najpierw ładowane do hurtowni "tak jak są", a dopiero potem mogą być przekształcane i analizowane zgodnie z wymaganiami. Zapewnia to większą elastyczność i możliwość przechowywania heterogenicznych danych, w tym danych nieustrukturyzowanych, takich jak tekst, obrazy i dzienniki.



Rys. 8.1-7 W przeciwieństwie do ETL, Data Lake wykorzystuje ELT, w którym informacje są najpierw przesyłane w "surowej" formie, a transformacja jest wykonywana na etapie przesyłania.

Tradycyjne hurtownie danych koncentrują się na wstępny przetwarzaniu danych w celu zapewnienia wysokiej wydajności zapytań, podczas gdy jeziora danych priorytetowo traktują elastyczność: przechowują surowe dane i przekształcają je w razie potrzeby (Rys. 8.1-8).



Rys. 8.1-8 Nowoczesne koncepcje pamięci masowej mają na celu przechowywanie i przetwarzanie wszystkich rodzajów danych do celów decyzyjnych.

Jednak pomimo wszystkich zalet, jeziora danych nie są pozbawione wad. Brak ścisłej struktury i złożoność

zarządzania informacjami może prowadzić do chaosu, w którym dane są powielane, sprzeczne ze sobą lub stają się nieistotne. Ponadto wyszukiwanie i analizowanie danych w takim repozytorium wymaga znacznego wysiłku, zwłaszcza gdy mamy do czynienia z heterogenicznymi informacjami. Aby przezwyciężyć te ograniczenia i połączyć najlepsze cechy tradycyjnych hurtowni danych i jezior danych, opracowano architekturę Data Lakehouse.

Architektura Data Lakehouse: synergia hurtowni i jezior danych

Aby połączyć najlepsze cechy DWH (ustrukturyzowana, zarządzalna, wysokowydajna analityka) i Data Lake (skalowalność, obsługa heterogenicznych danych), opracowano podejście Data Lakehouse. Architektura ta łączy elastyczność jezior danych z potężnymi narzędziami do przetwarzania i zarządzania typowymi dla tradycyjnych hurtowni, zapewniając równowagę między przechowywaniem, analizą i uczeniem maszynowym. Data Lakehouse to synteza jezior danych i hurtowni danych, łącząca elastyczność i skalowalność tych pierwszych z łatwością zarządzania i optymalizacją zapytań tych drugich.

Data Lakehouse to podejście architektoniczne, które ma na celu połączenie elastyczności i skalowalności jezior danych z łatwością zarządzania i wydajnością zapytań hurtowni danych (Rys. 8.1-9).

Kluczowe funkcje Data Lakehouse obejmują:

- **Otwarty format przechowywania danych:** korzystanie z otwartych formatów przechowywania danych, takich jak Apache Parquet, zapewnia wydajność i optymalizację zapytań.
- **Schemat tylko do odczytu:** w przeciwieństwie do tradycyjnego podejścia schematu tylko do zapisu w DWH, Lakehouse obsługuje schemat tylko do odczytu, co pozwala na większą elastyczność w zarządzaniu strukturą danych.
- **Elastyczność i skalowalność:** obsługuje przechowywanie i analizę ustrukturyzowanych i nieustrukturyzowanych danych, zapewniając wysoką wydajność zapytań dzięki optymalizacji na poziomie pamięci masowej.

Data Lakehouse oferuje kompromisowe rozwiązanie, które łączy w sobie zalety obu podejść, dzięki czemu idealnie nadaje się do nowoczesnych obciążeń analitycznych, które wymagają elastyczności w przetwarzaniu danych.



Rys. 8.1-9 Data Lakehouse to następna generacja systemów pamięci masowej zaprojektowanych w celu spełnienia złożonych i stale zmieniających się wymagań.

Idea stojąca za nowoczesnymi hurtowniami danych wydaje się prosta: jeśli wszystkie dane znajdują się w jednym miejscu, łatwiej jest je analizować. Jednak w praktyce nie wszystko przebiega tak gładko. Wyobraźmy sobie, że firma decyduje się całkowicie zrezygnować ze zwykłych systemów księgowych i zarządzania (ERP, PMIS, CAFM lub innych), zastępując je jednym ogromnym jeziorem danych, do którego każdy ma dostęp. Co się stanie? Najprawdopodobniej powstanie chaos: dane będą powielane, sprzeczne, a krytyczne informacje zostaną utracone lub uszkodzone. Nawet jeśli jezioro danych jest wykorzystywane wyłącznie do celów analitycznych, bez odpowiedniego zarządzania będzie ono poważnie zagrożone:

- Dane są trudne do zrozumienia: w konwencjonalnych systemach dane mają jasną strukturę, ale w jeziorze jest to po prostu ogromne nagromadzenie plików i tabel. Aby coś znaleźć, musi dowiedzieć się, za co odpowiada każdy wiersz i kolumna.
- Dane mogą być niedokładne: jeśli wiele wersji tych samych informacji jest przechowywanych w jednym miejscu, trudno jest ustalić, która wersja jest aktualna. W rezultacie decyzje podejmowane są na podstawie nieaktualnych lub błędnych danych.
- Przygotowanie danych do pracy jest trudne: dane muszą być nie tylko przechowywane, ale także prezentowane w wygodnej formie - w postaci raportów, wykresów, tabel. W tradycyjnych systemach odbywa się to automatycznie, ale w jeziorach danych wymaga dodatkowego przetwarzania.

W rezultacie każda koncepcja hurtowni danych ma swoje własne cechy, podejścia do przetwarzania i zastosowania biznesowe. Tradycyjne bazy danych koncentrują się na operacjach transakcyjnych, hurtownie danych (DWH) zapewniają strukturę dla analityki, jeziora danych (Data Lake) przechowują informacje w postaci nieprzetworzonej, a hurtownie hybrydowe (Data Lakehouse) łączą zalety DWH i Data Lake (Rys. 8.1-10).

	Traditional Approach	Data Warehouse	Data Lake	Data Lakehouse
Data Types	Relational Databases	Structured, ready for analytics	Raw, semi-structured, or unstructured	Mix of structured and unstructured
Use Cases	Transactional Systems	Reporting, dashboards, BI	Big data storage, AI, advanced analytics	Hybrid analytics, AI, real-time data
Processing	OLTP – real-time transactions	ETL – clean and structure before analysis	ELT – store raw data, transform later	ELT with optimized storage and real-time processing
Storage	On-premise servers	Centralized, SQL-based	Decentralized, flexible formats	Combines advantages of DWH and DL
Common Tools	MySQL, PostgreSQL	Snowflake, Redshift, BigQuery	Hadoop, AWS S3, Azure Data Lake	Databricks, Snowflake, Google BigLake

Rys. 8.1-10 DWH, Data Lake i Data Lakehouse: kluczowe różnice w typach danych, scenariuszach użycia, metodach przetwarzania i podejściach do przechowywania.

Wybór architektury pamięci masowej jest złożonym procesem, zależnym od potrzeb biznesowych, ilości informacji i wymagań analitycznych. Każde rozwiązanie ma swoje wady i zalety: DWH zapewnia strukturę, Data Lake zapewnia elastyczność, a Lakehouse zapewnia równowagę między tymi dwoma rozwiązaniami. Organizacje rzadko ograniczają się do jednej architektury danych.

Niezależnie od wybranej architektury, zautomatyzowane systemy zarządzania danymi znacznie przewyższają metody ręczne. Minimalizują błędy ludzkie, przyspieszają przetwarzanie informacji oraz zapewniają przejrzystość i identyfikowalność danych na wszystkich etapach procesów biznesowych.

Podczas gdy scentralizowane hurtownie danych stały się już standardem branżowym w wielu obszarach gospodarki, sytuacja w budownictwie pozostaje fragmentarna. Dane tutaj są rozproszone na różnych platformach (CDE, PMIS, ERP itp.), co utrudnia stworzenie jednolitego obrazu tego, co się dzieje i wymaga architektur zdolnych do łączenia tych źródeł w holistyczne, analitycznie użyteczne środowisko cyfrowe.

CDE, PMIS, ERP lub DWH i Data Lake

Niektóre firmy budowlane i inżynierijne wykorzystują już koncepcję Wspólnego Środowiska Danych (CDE) zgodnie z normą ISO 19650. Zasadniczo CDE pełni te same funkcje, co hurtownia danych (DWH) w innych branżach: centralizacja informacji, zapewnienie kontroli wersji, zapewnienie dostępu do zweryfikowanych informacji.

Common Data Environment (CDE) to scentralizowana przestrzeń cyfrowa wykorzystywana do zarządzania, przechowywania, udostępniania i współpracy nad informacjami projektowymi na wszystkich etapach cyklu życia obiektu. CDE jest często wdrażane przy użyciu technologii opartych na chmurze i zintegrowane z systemami CAD (BIM).

Sektor finansowy, detaliczny, logistyczny i przemysłowy od dziesięcioleci korzystają ze skoncentrowanych systemów zarządzania danymi, łącząc informacje z różnych źródeł, kontrolując ich znaczenie i zapewniając analitykę. CDE rozwija te zasady, dostosowując je do wyzwań związanych z projektowaniem budynków i zarządzaniem cyklem życia.

Podobnie jak DWH, CDE strukturyzuje dane, rejestruje zmiany i zapewnia pojedynczy punkt dostępu do zweryfikowanych informacji. Wraz z przejściem do chmury i integracją z narzędziami analitycznymi, różnice między nimi stają się coraz mniej widoczne. Dodając do CDE dane granularne, których koncepcja jest omawiana przez dostawców CAD od 2023 roku[93, 125], można dostrzec jeszcze więcej podobieństw do klasycznego DWH.

Wcześniej w rozdziale "Systemy ERP dla budownictwa i PMIS" przyjrzaliśmy się już systemom PMIS (Project Management Information System) i ERP (Enterprise Resource Planning). W projektach budowlanych CDE i PMIS współpracują ze sobą: CDE służy jako repozytorium danych, w tym rysunków, modeli i dokumentacji projektowej, podczas gdy PMIS zarządza procesami, takimi jak kontrolowanie terminów, zadań, zasobów i budżetów.

ERP, odpowiedzialny za zarządzanie firmą jako całością (finanse, zaopatrzenie, personel, produkcja), może integrować się z PMIS, zapewniając kontrolę kosztów i budżetu na poziomie firmy. Do celów analitycznych i raportowania, DWH może być wykorzystywany do gromadzenia, strukturyzowania i agregowania danych z CDE, PMIS i ERP w celu oceny finansowych wskaźników KPI (ROI) i identyfikacji wzorców. Z kolei Data Lake (DL) może uzupełniać DWH poprzez przechowywanie surowych i nieustrukturyzowanych danych (np. logów, danych z czujników, obrazów). Dane te mogą być przetwarzane i ładowane do DWH w celu dalszej analizy.

Tak więc CDE i PMIS koncentrują się na zarządzaniu projektami, ERP koncentruje się na procesach biznesowych, a DWH i Data Lake koncentrują się na analityce i operacjach na danych.

Porównując systemy CDE, PMIS i ERP z DWH i Data Lake, można zauważać znaczące różnice w zakresie niezależności od dostawcy, kosztów, elastyczności integracji, niezależności danych, szybkości adaptacji do zmian i możliwości analitycznych (Rys. 8.1-11). Tradycyjne systemy, takie jak CDE, PMIS i ERP, są często powiązane z konkretnymi rozwiązaniami i standardami dostawców, co czyni je mniej elastycznymi i zwiększa ich koszty ze względu na licencje i wsparcie. Ponadto dane w takich systemach są często zamknięte w zastrzeżonych, zamkniętych formatach, co ogranicza ich wykorzystanie i analizę.

		CDE, PMIS, ERP	DWH, Data Lake
	Vendor Dependency	High (tied to specific solutions and standards of vendors)	Low (flexibility in tool and platform choice)
	Integration Flexibility	Limited (integration depends on vendor solutions)	High (easily integrates with various data sources)
	Cost	High (licensing and support costs)	Relatively lower (use of open technologies and platforms)
	Data Independence	Low (data often locked in proprietary formats)	High (data stored in open and accessible formats)
	Adaptability to Changes	Slow (changes require vendor approval and integration)	Fast (adaptation and data structure modification without intermediaries)
	Analytical Capabilities	Limited (dependent on vendor-provided solutions)	Extensive (support for a wide range of analytical tools)

Rys. 8.1-11 DWH i Data Lake oferują większą elastyczność i niezależność danych niż systemy takie jak CDE, PMIS i ERP.

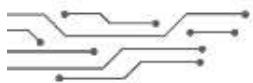
Z kolei DWH i Data Lake zapewniają większą elastyczność w integracji z różnymi źródłami danych, a ich wykorzystanie otwartych technologii i platform pomaga obniżyć całkowity koszt posiadania. Co więcej, DWH i Data Lake obsługują szeroką gamę narzędzi analitycznych, co zwiększa możliwości analizy i zarządzania.

Wraz z rozwojem narzędzi do inżynierii odwrotnej dla formatów CAD i dostępem do baz danych aplikacji CAD, pytanie staje się coraz bardziej dotkliwe: jak uzasadnione jest dalsze korzystanie z zamkniętych, odizolowanych platform, jeśli dane projektowe muszą być dostępne dla szerokiego grona specjalistów pracujących w dziesiątkach wykonawców i organizacji projektowych?

Ta zależność od technologii konkretnego dostawcy może znacznie ograniczyć elastyczność zarządzania danymi, spowolnić reakcję na zmiany w projekcie i utrudnić skuteczną współpracę między uczestnikami.

Tradycyjne podejścia do zarządzania danymi - w tym DWH, Data Lake, CDE i PMIS - koncentrowały się przede wszystkim na przechowywaniu, strukturyzowaniu i przetwarzaniu informacji. Jednak wraz z rozwojem sztucznej inteligencji i uczenia maszynowego rośnie zapotrzebowanie na nowe sposoby organizacji danych, które nie tylko agregują, ale także ujawniają złożone relacje, znajdują ukryte wzorce i zapewniają natychmiastowy dostęp do najbardziej istotnych informacji.

Szczególną rolę w tym kierunku zaczynają odgrywać wektorowe bazy danych - nowy typ pamięci masowej zoptymalizowany pod kątem wielowymiarowego osadzania.



ROZDZIAŁ 7.2.

ZARZĄDZANIE HURTOWNIĄ DANYCH I ZAPOBIEGANIE CHAOSOWI

Wektorowe bazy danych i pole ograniczające

Wektorowe bazy danych to nowa klasa repozytoriów, które nie tylko przechowują dane, ale umożliwiają wyszukiwanie według znaczenia, porównywanie obiektów według semantycznej bliskości i tworzenie inteligentnych systemów: od rekomendacji po automatyczną analizę i generowanie kontekstu. W przeciwieństwie do tradycyjnych baz danych, które koncentrują się na dokładnych dopasowaniach, wektorowe bazy danych znajdują podobne obiekty na podstawie atrybutów - nawet jeśli nie ma dokładnego dopasowania

Wektorowa baza danych jest wyspecjalizowanym typem bazy danych, która przechowuje dane jako wielowymiarowe wektory, z których każdy reprezentuje pewne cechy lub właściwości.

Wektory te mogą mieć różną liczbę wymiarów, w zależności od złożoności danych (w jednym przypadku może to być kilka wymiarów, a w innym — tysiące).

Główną zaletą wektorowych baz danych jest wyszukiwanie według znaczenia semantycznego, a nie dokładnego dopasowania wartości. Zamiast zapytań SQL i Pandas z filtrami "równa się" lub "zawiera", używane jest wyszukiwanie najbliższych sąsiadów (k-NN) (więcej o k-NN powiemy w następnej części książki) w przestrzeni cech.

Wraz z rozwojem LLM (Large Language Models) i modeli generatywnych, interakcja z bazami danych zaczyna się zmieniać. Obecnie możliwe jest wyszukiwanie danych w języku naturalnym, semantyczne przeszukiwanie dokumentów, automatyczne wyodrębnianie kluczowych terminów i budowanie kontekstowych relacji między obiektami - wszystko to bez konieczności biegłej znajomości SQL lub znajomości struktury tabel. Zostało to omówione bardziej szczegółowo w sekcji "LLM i ich rola w przetwarzaniu danych i procesach biznesowych".

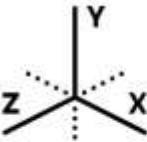
Ważne jest jednak, aby zdawać sobie sprawę, że LLM nie strukturyzują i nie porządkują informacji automatycznie. Model po prostu przepływa przez dane i znajduje najbardziej odpowiedni fragment danych w oparciu o kontekst zapytania. Jeśli dane nie zostały wstępnie oczyszczone lub przekształcone, głębokie wyszukiwanie będzie jak próba znalezienia odpowiedzi w cyfrowych "śmieciach" - może zadziałać, ale jakość wyników będzie niższa. Idealnie, jeśli dane można ustukturyzować (np. przetłumaczyć dokumenty na Markdown) i załadować do wektorowej bazy danych. To znacznie zwiększa dokładność i trafność wyników.

Początkowo wektorowe bazy danych były wykorzystywane w uczeniu maszynowym, ale obecnie znajdują coraz więcej zastosowań poza nim - w wyszukiwarkach, personalizacji treści i intelligentnej analityce.

Jednym z najbardziej oczywistych przykładów podejścia wektorowego w konstrukcji jest Bounding Box (równoległy ograniczający). Jest to konstrukcja geometryczna opisująca granice obiektu w przestrzeni trójwymiarowej. Bounding Box jest definiowany przez minimalne i maksymalne współrzędne X,

Y i Z, tworząc "pułapkę" wokół obiektu. Metoda ta pozwala oszacować rozmiar i umiejscowienie elementu bez konieczności analizowania całej geometrii.

Każde Box może być reprezentowane jako wektor w przestrzeni wielowymiarowej: na przykład [x, y, z, szerokość, wysokość, głębokość] - już 6 wymiarów (rys. 8.2-1).



Bounding Box

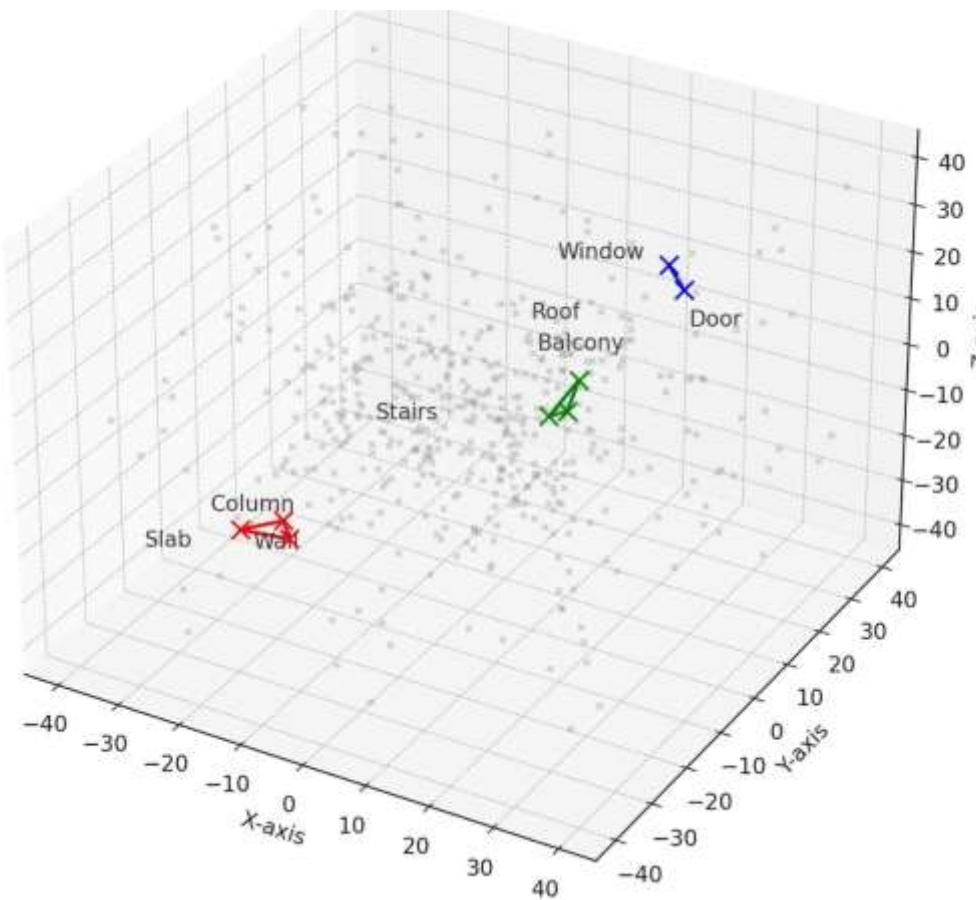
	minX	maxX	minY	maxY	minZ	maxZ	Width	Height	Depth
Column	-15	-5	-25	-15	0	10	10	10	20
Stairs	-5	5	-15	-5	0	10	10	10	10
Door	5	15	5	15	0	10	10	10	10
Window	25	35	-35	-25	10	30	10	20	20
Balcony	15	25	-5	5	20	40	10	20	20

Rys. 8.2-1 Bounding Box - informacje o współrzędnych elementów i ich lokalizacji w modelu projektu są analogiczne do wektorowej bazy danych.

Taka reprezentacja danych ułatwia wiele zadań, w tym sprawdzanie przecięć między obiekta mi, planowanie przestrzennego rozmieszczenia elementów budynku i wykonywanie zautomatyzowanych obliczeń. Bounding Box może służyć jako mostek między złożonymi modelami 3D a tradycyjnymi wektorowymi bazami danych, umożliwiając efektywne wykorzystanie zalet obu podejść w modelowaniu architektonicznym i inżynieryjnym.

Bounding Box to "wektoryzacja geometrii", a osadzanie (sposób przekształcania czegoś abstrakcyjnego) to "wektoryzacja znaczenia". Oba podejścia pozwalają przejść od ręcznego wyszukiwania do intelligentnego wyszukiwania, czy to obiektów 3D w modelu projektu, czy pojęć w tekście.

Wyszukiwanie obiektów w projekcie (na przykład "znajdź wszystkie okna o szerokości > 1,5 m") jest podobne do wyszukiwania najbliższych sąsiadów (k-NN) w wektorowej bazie danych, gdzie kryteria definiują "strefę" w przestrzeni cech. (więcej o wyszukiwaniu najbliższych sąsiadów k-NN powiemy w następnej części dotyczącej uczenia maszynowego) (rys. 8.2-2). Jeśli dodamy dodatkowe parametry (materiał, waga, czas produkcji) do atrybutów obwiedni, tabela zmieni się w wektor wielowymiarowy, w którym każdy atrybut jest nowym wymiarem. Jest to bliższe nowoczesnym bazom wektorowym, w których wymiary liczone są w setkach lub tysiącach (np. osadzanie z sieci neuronowych).



Rys. 8.2-2 Wyszukiwanie obiektów w projekcie przy użyciu wektorowych baz danych.

Podejście zastosowane w Bounding Box, ma zastosowanie nie tylko do obiektów geometrycznych, ale także do analizy tekstu i języka. Wektorowe reprezentacje danych są już aktywnie wykorzystywane w przetwarzaniu języka naturalnego (NLP). Podobnie jak obiekty w projekcie budowlanym mogą być grupowane według ich bliskości przestrzennej (Rys. 8.2-2), słowa w tekście mogą być analizowane według ich semantycznej i kontekstowej bliskości.

Na przykład słowa "architekt", "konstrukcja", "projekt" będą znajdować się obok siebie w przestrzeni wektorowej, ponieważ mają podobne znaczenie. W LLM mechanizm ten umożliwia automatyczną, niewymagającą ręcznej kategoryzacji:

- Identyfikacja tematu tekstu
- Przeprowadzanie wyszukiwania semantycznego w treści dokumentów
- Generowanie automatycznych adnotacji i podsumowań tekstu
- Znajdź synonimy i terminy pokrewne

Wektorowe bazy danych pozwalają analizować tekst i znajdować w nim powiązane terminy w taki sam sposób, w jaki Bounding Box pomaga analizować obiekty przestrzenne w modelach 3D. Przykład Bounding Box elementów projektu pomaga zrozumieć, że reprezentacja wektorowa nie jest czysto "sztuczną" koncepcją ML, ale naturalnym sposobem strukturyzacji danych w celu rozwiązywania problemów

użytkowych, niezależnie od tego, czy jest to wyszukiwanie kolumn w projekcie CAD, czy semantycznie bliskich obrazów w bazie danych.

Specjaliści pracujący z bazami danych powinni zwrócić uwagę na magazyny wektorowe. Ich rozprzestrzenianie się wskazuje na nowy etap w rozwoju baz danych, w którym klasyczne systemy relacyjne i technologie zorientowane na AI zaczynają się przeplatać, tworząc hybrydowe rozwiązania przyszłości.

Użytkownicy opracowujący złożone i wielkoskalowe aplikacje AI będą korzystać ze specjalistycznych baz danych do wyszukiwania wektorowego. Jednocześnie ci, którzy potrzebują tylko oddzielnych funkcji sztucznej inteligencji do integracji z istniejącymi aplikacjami, częściej wybierają wbudowane funkcje wyszukiwania wektorowego w bazach danych, z których już korzystają (PostgreSQL, Redis).

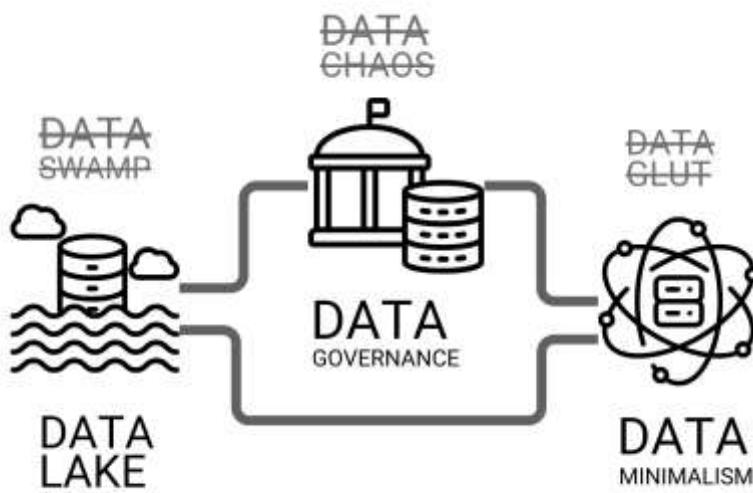
Chociaż systemy takie jak DWH, Data Lake, CDE, PMIS, wektorowe bazy danych i inne oferują różne podejścia do przechowywania danych i zarządzania nimi, ich skuteczność zależy nie tylko od ich architektury, ale także od tego, jak dobrze same dane są zorganizowane i zarządzane. Nawet w przypadku korzystania z nowoczesnych rozwiązań - czy to wektorowych baz danych, klasycznych relacyjnych DBMS czy hurtowni typu Data Lake - brak jasnych zasad zarządzania, strukturyzacji i aktualizacji danych może prowadzić do tych samych trudności, z którymi borykają się użytkownicy pracujący z różnymi plikami i danymi wieloformatowymi.

Bez Data Governance), nawet najpotężniejsze rozwiązania mogą stać się chaotyczne i niestrukturyzowane, zamieniając jeziora danych w bagna danych). Aby tego uniknąć, firmy muszą nie tylko wybrać odpowiednią architekturę pamięci masowej, ale także wdrożyć minimalizm danych), zarządzanie dostępem i strategie kontroli jakości, aby przekształcić dane w skuteczne narzędzie do podejmowania decyzji.

Data Governance), Data Minimalism) i Data Swamp)

Zrozumienie i wdrożenie koncepcji Data Governance), Data Minimalism) i zapobieganie Data Swamp) są kluczem do skutecznego zarządzania hurtowniami danych i dostarczania wartości biznesowej (Rys. 8.2-3).

Według badania przeprowadzonego przez firmę Gartner (2017), 85% projektów Big Data kończy się niepowodzeniem, a jednym z kluczowych powodów jest niewystarczająca jakość danych i zarządzanie danymi [144].



Rys. 8.2-3 Niektóre z kluczowych aspektów zarządzania danymi to Data Governance i Data Minimalism.

Zarządzanie danymi (Data Governance) jest podstawowym elementem zarządzania danymi, zapewniającym odpowiednie i skuteczne wykorzystanie danych we wszystkich procesach biznesowych. Chodzi nie tylko o ustanowienie zasad i procedur, ale także o zapewnienie dostępności, niezawodności i bezpieczeństwa danych:

- Definiowanie i klasyfikowanie danych: jasne definiowanie i klasyfikowanie podmiotów pozwala organizacjom zrozumieć, jakie podmioty są potrzebne w firmie i określić, w jaki sposób powinny być one wykorzystywane.
- Prawa dostępu i zarządzanie: opracowanie zasad i procedur dostępu do danych i zarządzania nimi zapewnia, że tylko upoważnieni użytkownicy mają dostęp do określonych danych.
- Ochrona danych przed zagrożeniami zewnętrznymi: Ochrona danych przed zagrożeniami zewnętrznymi jest kluczowym aspektem zarządzania danymi. Obejmuje to nie tylko środki techniczne, ale także szkolenie pracowników w zakresie podstaw bezpieczeństwa informacji.

Minimalizm danych (Data Minimalism) to podejście polegające na ograniczeniu danych do najbardziej wartościowych i znaczących atrybutów i jednostek w formacji (Rys. 8.2-4), zmniejszając w ten sposób koszty i poprawiając wykorzystanie danych:

- Uproszczenie procesu podejmowania decyzji: zmniejszenie liczby obiektów i ich atrybutów do tych najbardziej istotnych upraszcza proces podejmowania decyzji poprzez skrócenie czasu i zasobów wymaganych do analizy i przetwarzania danych.
- Skupienie się na tym, co ważne: wybór najbardziej istotnych jednostek i atrybutów pozwala skupić się na informacjach, które naprawdę mają znaczenie dla firmy, eliminując szum i niepotrzebne dane.
- Wydajna alokacja zasobów: minimalizacja danych umożliwia bardziej wydajną alokację zasobów, zmniejszając koszty przechowywania i przetwarzania danych, poprawiając jakość i bezpieczeństwo danych.

Logika pracy z danymi nie powinna rozpoczynać się od ich tworzenia jako takiego (rys. 8.2-4), ale od zrozumienia przyszłych scenariuszy wykorzystania tych danych jeszcze przed rozpoczęciem procesu

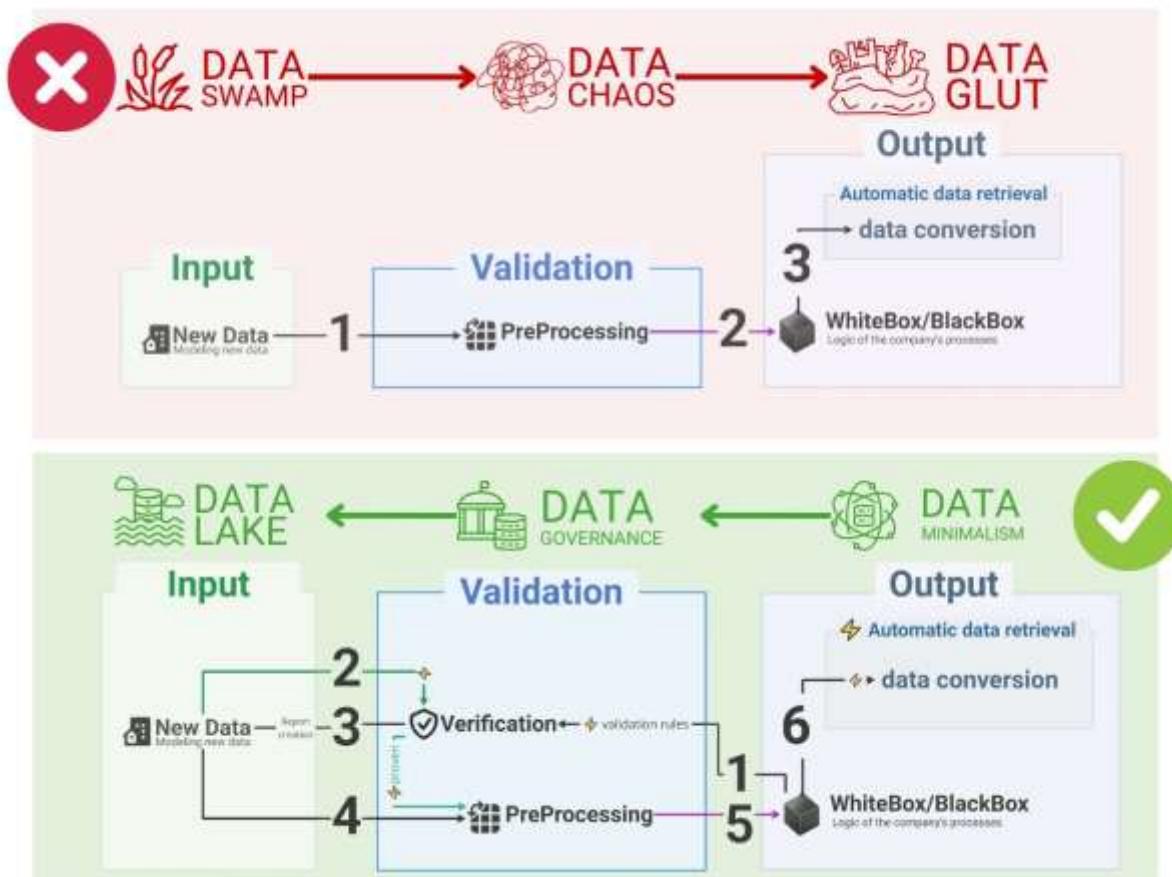
generowania. Takie podejście pozwala z góry zdefiniować minimalne niezbędne wymagania dotyczące atrybutów, ich typów i wartości granicznych. Wymagania te stanowią podstawę do tworzenia poprawnych i stabilnych encji w modelu informacyjnym. Wstępne zrozumienie celów i zastosowań danych przyczynia się do utworzenia struktury odpowiedniej do analizy. Więcej szczegółów na temat podejść do modelowania danych na poziomie koncepcyjnym, logicznym i fizycznym omówiono w rozdziale "Modelowanie danych: model koncepcyjny, logiczny i fizyczny".

W tradycyjnych procesach biznesowych firm budowlanych przetwarzanie danych częściej przypomina wrzucanie danych do bagna, gdzie najpierw dane są tworzone, a następnie specjalisi próbują je zintegrować z innymi systemami i narzędziami.

Bagno danych (Data Swamp) jest wynikiem niekontrolowanego gromadzenia i przechowywania danych bez odpowiedniej organizacji, strukturyzacji i zarządzania, w wyniku czego dane są niestrukturyzowane, trudne do wykorzystania i mają niewielką wartość.

Jak zapobiec sytuacji, w której przepływ informacji staje się bagnem:

- **Zarządzanie strukturą danych:** zapewnienie, że dane są uporządkowane i skategoryzowane, pomaga zapobiegać zalewovi danych, czyniąc je uporządkowanymi i łatwo dostępnymi.
- **Zrozumienie i interpretacja** danych: jasny opis pochodzenia, modyfikacji i znaczenia danych zapewnia ich prawidłowe zrozumienie i interpretację.
- **Utrzymanie jakości danych:** regularna konserwacja i czyszczenie danych pomaga utrzymać ich jakość, przydatność i wartość dla analityki i procesów biznesowych.



Rys. 8.2-4 Aby uniknąć bałaganu w hurtowni danych, należy rozpocząć proces tworzenia danych od zebrania wymagań dotyczących atrybutów.

Włączając zasady zarządzania danymi i minimalizmu danych do procesów zarządzania danymi oraz aktywnie zapobiegając przekształcaniu się hurtowni danych w bagna danych, organizacje mogą zmaksymalizować potencjał swoich danych.

Kolejnym etapem ewolucji pracy z danymi, po rozwiązaniu kwestii zarządzania i minimalizmu, jest standaryzacja automatycznego przetwarzania, zapewnienie jakości i wdrożenie metod, które sprawiają, że dane są użyteczne do analizy, transformacji i podejmowania decyzji. Tym właśnie zajmują się metodologie DataOps i VectorOps, które stają się ważnymi narzędziami dla firm pracujących z big data i uczeniem maszynowym.

DataOps i VectorOps: nowe standardy danych

Podczas gdy Data Governance odpowiada za kontrolę i organizację danych, DataOps pomaga zapewnić ich dokładność, spójność i płynny przepływ w firmie. Jest to szczególnie ważne w wielu przypadkach biznesowych w budownictwie, gdzie dane są generowane w sposób ciągły i wymagają terminowego przetwarzania. Na przykład w sytuacjach, w których modele informacji o budynku, wymagania projektu i raporty analityczne muszą być zsynchronizowane między różnymi systemami w ciągu jednego dnia roboczego, rolą

DataOps może być kluczowa. Umożliwia ona budowanie stabilnych i powtarzalnych procesów przetwarzania danych, zmniejszając ryzyko opóźnień i utraty przydatności informacji.

Samo zarządzanie danymi nie wystarczy - ważne jest, aby dane nie były tylko przechowywane, ale aktywnie wykorzystywane w codziennych operacjach. W tym miejscu do gry wkracza DataOps - metodologia skoncentrowana na automatyzacji, integracji i ciągłym przepływie danych.

DataOps koncentruje się na poprawie współpracy, integracji i automatyzacji przepływów danych w organizacjach. Przyjęcie praktyk DataOps promuje dokładność, spójność i dostępność danych, co ma kluczowe znaczenie dla aplikacji skoncentrowanych na danych.

Kluczowymi narzędziami w ekosystemie DataOps są Apache Airflow (Rys. 7.4-4) do orkiestracji przepływu pracy oraz Apache NiFi (Rys. 7.4-5) do routingu i przekształcania przepływów danych. Razem technologie te umożliwiają elastyczne, niezawodne i skalowalne potoki danych do automatycznego przetwarzania, kontroli i integracji informacji między systemami (więcej szczegółów w rozdziale "Automatyczny ETL -conveyor"). Wdrażając podejście DataOps w procesach budowlanych, należy wziąć pod uwagę cztery podstawowe aspekty:

1. **Ludzie i narzędzia są ważniejsi niż dane:** silosowe repozytoria danych mogą być postrzegane jako główny problem, ale rzeczywistość jest bardziej złożona. Oprócz fragmentacji danych, istotną rolę odgrywa izolacja zespołów i różne narzędzia, z których korzystają. W budownictwie z danymi pracują specjaliści z różnych dziedzin: inżynierowie danych i analitycy, zespoły BI i wizualizacji, a także eksperci ds. zarządzania projektami i jakości. Każdy z nich ma inny sposób pracy, dlatego ważne staje się stworzenie ekosystemu, w którym dane przepływają swobodnie między uczestnikami, zapewniając jedną, spijną wersję informacji.
2. **Automatyzacja testowania i wykrywania błędów:** Dane budowlane zawsze zawierają błędy, czy to niedokładności w modelach, błędy obliczeniowe czy nieaktualne specyfikacje. Regularne testowanie danych i eliminowanie powtarzających się błędów może znacznie poprawić jakość danych. W ramach DataOps należy wdrożyć zautomatyzowane mechanizmy kontroli i walidacji, które monitorują poprawność danych, analizują błędy i identyfikują wzorce oraz wychwytyują i usuwają awarie systemu w każdym przepływie pracy. Im wyższy stopień zautomatyzowanej walidacji, tym wyższa ogólna jakość danych i niższe prawdopodobieństwo wystąpienia błędów na końcowych etapach.
3. **Dane powinny być testowane w taki sam sposób, jak kod programu:** większość aplikacji budowlanych opiera się na przetwarzaniu danych, ale ich kontrola jest często pozostawiana drugorzędnym roliom. Jeśli modele uczenia maszynowego są szkolone na niedokładnych danych, prowadzi to do błędnych prognoz i strat finansowych. W ramach DataOps dane powinny być poddawane takiej samej rygorystycznej kontroli jak kod oprogramowania: kontrole logiczne, testy warunków skrajnych, ocena zachowania modeli w przypadku zmiany wartości wejściowych. Tylko zweryfikowane i wiarygodne dane mogą być wykorzystywane jako podstawa do podejmowania decyzji zarządczych.
4. **Obserwonalność danych bez uszczerbku dla wydajności:** monitorowanie danych nie jest tylko zbiorem metryk, ale strategicznym narzędziem zarządzania jakością. Aby DataOps działało skutecznie, obserwonalność musi być wbudowana na wszystkich etapach obsługi danych, od projektu do działania. Jednocześnie ważne jest, aby monitorowanie nie spowalniało systemu. W projektach budowlanych kluczowe jest nie tylko gromadzenie danych, ale także robienie tego w taki sposób, aby praca specjalistów (np. projektantów) tworzących dane nie była w żaden sposób zakłócona. Taka równowaga pozwala kontrolować jakość danych bez uszczerbku dla produktywności.

DataOps nie jest dodatkowym obciążeniem dla analityków danych, ale podstawą ich pracy. Wdrażając DataOps, firmy budowlane mogą przejść od chaotycznego zarządzania danymi do wydajnego ekosystemu, w którym dane pracują dla biznesu.

Z kolei VectorOps stanowi kolejny etap ewolucji DataOps, skupiając się na przetwarzaniu, przechowywaniu i analizowaniu wielowymiarowych danych wektorowych (które zostały omówione w poprzednim rozdziale). Jest to szczególnie istotne w obszarach takich jak cyfrowe bliźniaki, modele sieci neuronowych i wyszukiwanie semantyczne, które zaczynają pojawiać się w branży budowlanej. VectorOps opiera się na wektorowych bazach danych w celu wydajnego przechowywania, indeksowania i wyszukiwania wielowymiarowych reprezentacji obiektów.

VectorOps to kolejny krok po DataOps, koncentrujący się na przetwarzaniu, analizowaniu i wykorzystywaniu danych wektorowych w budownictwie. W przeciwieństwie do DataOps, który koncentruje się na przepływie danych, spójności i jakości, VectorOps koncentruje się na zarządzaniu wielowymiarowymi reprezentacjami obiektów wymaganymi do uczenia maszynowego.

W przeciwieństwie do tradycyjnych podejść, VectorOps pozwala uzyskać dokładniejsze opisy obiektów, co ma kluczowe znaczenie dla cyfrowych bliźniaków, systemów projektowania generatywnego i automatycznego wykrywania błędów w danych CAD konwertowanych do formatu wektorowego. Połączona implementacja DataOps i VectorOps tworzy solidną podstawę dla skalowej, zautomatyzowanej pracy z dużymi ilościami informacji - od klasycznych tabel po bogate semantycznie modele przestrzenne

Kolejne kroki: od chaotycznej pamięci masowej do uporządkowanej pamięci masowej

Tradycyjne podejście do przechowywania danych często skutkuje tworzeniem różnych "silosów informacji", w których ważne informacje są niedostępne do analizy i podejmowania decyzji. Nowoczesne koncepcje przechowywania danych, takie jak Data Warehouse, Data Lake i ich hybrydy, umożliwiają ujednolicenie rozproszonych informacji i udostępnienie ich w skoncentrowany sposób na potrzeby strumieniowego przesyłania danych i analizy biznesowej. Ważne jest nie tylko wybranie odpowiedniej architektury pamięci masowej, ale także wdrożenie zarządzania danymi i minimalizmu danych), aby zapobiec przekształceniu się pamięci masowych w niekontrolowane bagna danych).

Podsumowując tę część, warto podkreślić główne praktyczne kroki, które pomogą ci zastosować omówione koncepcje w codziennych zadaniach:

- Wybór wydajnych formatów przechowywania danych
 - Odejście od CSV i XLSX na rzecz bardziej wydajnych formatów (Apache Parquet, ORC) do przechowywania dużych ilości danych.
 - Wdrożenie systemu wersjonowania danych w celu śledzenia zmian
 - Wykorzystanie metadanych do opisania struktury i pochodzenia informacji.
- Tworzenie ujednoliconej architektury danych firmy
 - Porównaj różne architektury pamięci masowej: RDBMS, DWH i Data Lake. Wybierz tą, która

najlepiej spełnia Twoje potrzeby w zakresie skalowalności, integracji źródeł i przetwarzania analitycznego.

- Zaprojektuj mapę procesu wyodrębniania, ładowania i przekształcania danych (ETL) z różnych źródeł dla swoich zadań. Użyj narzędzi do wizualizacji, takich jak Miro, Lucidchart lub Draw.io, aby zwizualizować kluczowe kroki i punkty integracji.

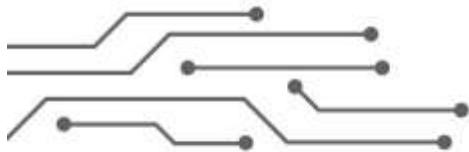
■ Wdrożenie praktyk Data Governance i Data Minimalism

- Postępuj zgodnie z podejściem Data Minimalism - przechowuj i przetwarzaj tylko to, co jest naprawdę wartościowe.
- Wdrożenie zasad zarządzania danymi - określenie odpowiedzialności za dane, zapewnienie jakości i przejrzystości.
- Dowiedz się więcej o zasadach zarządzania danymi i koncepcjach DataOps, VectorOps
- Definiowanie kryteriów jakości danych i procedur walidacji danych w ramach DataOps

Dobrze zorganizowane przechowywanie danych tworzy podstawę do centralizacji procesów analitycznych w firmie. Przejście od chaotycznego gromadzenia plików do uporządkowanego przechowywania pozwala przekształcić informacje w strategiczny zasób, który pomaga podejmować świadome decyzje i poprawiać wydajność procesów biznesowych.

Po zautomatyzowaniu i standaryzacji procesów gromadzenia, przekształcania, analizy i przechowywania danych, kolejnym etapem transformacji cyfrowej jest pełnoprawna obsługa Big Data.





VIII CZĘŚĆ

DUŻE ZBIORY DANYCH, UCZENIE MASZYNOWE I PROGNOZY

Dziewiąta część skupia się na dużych zbiorach danych, uczeniu maszynowym i analityce predykcyjnej w branży budowlanej. Bada przejście od intuicyjnego podejmowania decyzji do obiektywnych analiz opartych na danych historycznych. Praktyczne przykłady są wykorzystywane do zademonstrowania analizy dużych zbiorów danych w budownictwie - od analizowania zbioru danych pozwoleń na budowę w San Francisco po przetwarzanie CAD - projektów z milionami elementów. Szczególną uwagę poświęcono metodom uczenia maszynowego do przewidywania kosztów i harmonogramu projektów budowlanych, ze szczególnym omówieniem algorytmów regresji liniowej i k-najbliższych sąsiadów. Pokazano, w jaki sposób ustrukturyzowane dane stają się podstawą modeli predykcyjnych do oceny ryzyka, optymalizacji zasobów i poprawy efektywności zarządzania projektami. Część ta zawiera również zalecenia dotyczące wyboru reprezentatywnych próbek danych i wyjaśnia, dlaczego duże zbiory danych nie zawsze są wymagane do skutecznych analiz.

ROZDZIAŁ 8.1.

BIG DATA I ICH ANALIZA

Big Data w budownictwie: od intuicji do przewidywalności

Termin "big data" nie ma ścisłej definicji. Pojęcie to pojawiło się pierwotnie, gdy ilość informacji zaczęła przekraczać możliwości tradycyjnych metod ich przetwarzania. Obecnie ilość i złożoność danych w wielu branżach, w tym w budownictwie, wzrosła tak bardzo, że nie mieścią się one w lokalnej pamięci komputerów i wymagają zastosowania nowych technologii do ich przetwarzania.

Istotą pracy z Big Data jest nie tylko przechowywanie i przetwarzanie danych, ale także możliwości predykcyjne. W branży budowlanej Big Data otwiera drogę od intuicyjnych decyzji opartych na subiektywnej interpretacji tabel i wizualizacji (jak omówiono wcześniej) do świadomych prognoz popartych rzeczywistymi obserwacjami i statystykami.

Wbrew powszechnemu przekonaniu, celem pracy z dużymi zbiorami danych nie jest "sprawienie, by maszyna myślała jak człowiek", ale zastosowanie modeli matematycznych i algorytmów do analizy ogromnych ilości danych w celu identyfikacji wzorców, przewidywania zdarzeń i optymalizacji procesów.

Big Data nie jest zimnym światem algorytmów pozbawionym wpływu człowieka. Wręcz przeciwnie, big data działa w połączeniu z naszymi instynktami, błędami i kreatywnością. To właśnie niedoskonałość ludzkiego myślenia pozwala nam znajdować niestandardowe rozwiązania i dokonywać przełomów.

Wraz z rozwojem technologii cyfrowej, branża budowlana zaczęła aktywnie wykorzystywać techniki przetwarzania danych pochodzące z sektora IT. Dzięki narzędziom takim jak Pandas i Apache Parquet można łączyć dane ustrukturyzowane i nieustrukturyzowane, upraszczając dostęp do informacji i zmniejszając straty w analizie, podczas gdy duże zbiory danych z dokumentów lub projektów CAD (rys. 9.2-10 - rys. 9.2-12) umożliwiają gromadzenie, analizowanie i przewidywanie danych na wszystkich etapach cyklu życia projektu.

Big Data ma transformacyjny wpływ na branżę budowlaną, wpływając na nią potencjalnie na wiele sposobów. Zastosowanie technologii Big Data przynosi rezultaty w wielu kluczowych obszarach, w tym np:

- **Analiza potencjału inwestycyjnego** - prognozowanie rentowności i okresów zwrotu projektów na podstawie danych z poprzednich obiektów.
- **Konserwacja predykcyjna** - identyfikacja prawdopodobnych awarii sprzętu przed ich wystąpieniem, co skraca czas przestojów.
- **Optymalizacja łańcucha dostaw** - przewidywanie zakłóceń i poprawa wydajności logistyki.
- **Analizy efektywności energetycznej** - pomoc w projektowaniu budynków niskoenergetycznych.
- **Monitorowanie bezpieczeństwa** - wykorzystanie czujników i urządzeń do noszenia w celu monitorowania warunków w miejscu pracy.
- **Kontrola jakości** - monitorowanie w czasie rzeczywistym zgodności ze standardami procesu.

■ Zarządzanie personelem - analiza wydajności i prognozowanie zapotrzebowania na personel.

Trudno jest znaleźć obszar w budownictwie, w którym analiza danych i przewidywanie nie są pożądane. Główną zaletą algorytmów predykcyjnych jest ich zdolność do samouczenia się i ciągłego doskonalenia w miarę gromadzenia danych.

W niedalekiej przyszłości sztuczna inteligencja będzie nie tylko pomagać budowniczym, ale także podejmować kluczowe decyzje - od procesów projektowych po kwestie związane z eksploatacją budynku.

Więcej o tym, jak generowane są prognozy i wykorzystywane są modele uczenia się, zostanie omówione w następnej części książki, "Machine Learning and Predictions".

Przejście do pełnoprawnej pracy z big data wymaga zmiany samego podejścia do analityki. Podczas gdy klasyczne systemy, które rozważaliśmy do tej pory, koncentrowały się na związkach przyczynowo-skutkowych, analityka dużych zbiorów danych przenosi swój punkt ciężkości na poszukiwanie statystycznych wzorców i korelacji, które pozwalają nam identyfikować ukryte relacje i przewidywać zachowanie obiektów nawet bez pełnego zrozumienia wszystkich czynników.

Kwestionowanie wykonalności dużych zbiorów danych: korelacja, statystyka i próbkowanie danych

Tradycyjnie budownictwo opierało się na subiektywnych hipotezach i osobistym doświadczeniu. Inżynierowie zakładali - z pewnym prawdopodobieństwem - jak zachowa się materiał, jakie obciążenia wytrzyma konstrukcja i jak długo potrwa projekt. Założenia te były testowane w praktyce, często kosztem czasu, zasobów i przyszłych zagrożeń.

Wraz z pojawiением się dużych zbiorów danych, podejście to zmienia się diametralnie: decyzje nie są już podejmowane na podstawie intuicyjnych przeczuć, ale w wyniku analizy dużych zbiorów danych. Budownictwo stopniowo przestaje być sztuką intuicji i staje się precyzyjną nauką przewidywania.

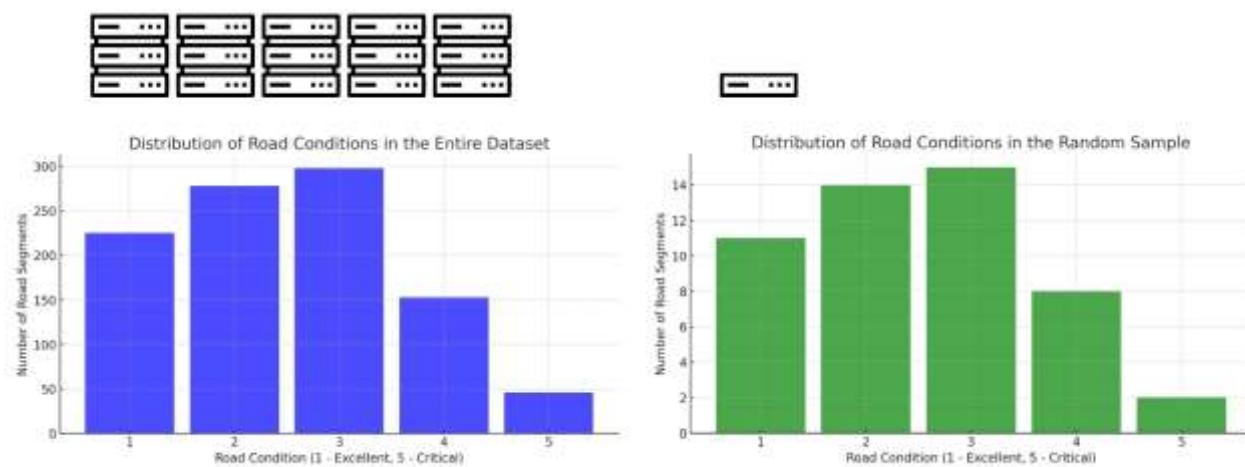
Przejście do idei korzystania z dużych zbiorów danych nieuchronnie rodzi ważne pytanie: jak krytyczna jest ilość danych i ile informacji jest naprawdę potrzebnych do niezawodnej analizy predykcyjnej? Powszechnie przekonanie, że "im więcej danych, tym większa dokładność" nie zawsze okazuje się statystycznie uzasadnione w praktyce.

Już w 1934 roku statystyk Jerzy Neumann udowodnił [145], że kluczem do dokładności wnioskowania statystycznego jest nie tyle ilość danych, co ich reprezentatywność i losowość doboru próby.

Jest to szczególnie prawdziwe w branży budowlanej, gdzie duże masy danych są gromadzone za pomocą IoT -czujników, skannerów, kamer monitorujących, dronów, a nawet wieloformatowych modeli CAD, co zwiększa ryzyko martwych punktów, wartości odstających i zniekształceń danych.

Rozważmy przykład monitorowania stanu nawierzchni dróg. Kompletny zestaw danych obejmujący wszystkie odcinki dróg może zajmować X GB, a jego przetworzenie może zająć około jednego dnia.

Jednocześnie losowa próbka obejmująca tylko co 50. odcinek drogi zajęłaby tylko $X/50$ GB i zostałaby przetworzona w pół godziny, zapewniając jednocześnie podobną dokładność szacunków dla niektórych obliczeń (Rys. 9.1-1).



Rys. 9.1-1 Histogramy stanu nawierzchni: pełny zestaw danych i losowe pobieranie próbek pokazują identyczne wyniki.

Tak więc kluczem do udanej analizy danych może być często nie ilość danych, ale reprezentatywność próby i jakość zastosowanych metod przetwarzania. Przejście na losowe pobieranie próbek i bardziej selektywne podejście wymaga zmiany sposobu myślenia w branży budowlanej. W przeszłości firmy kierowały się logiką "im więcej danych, tym lepiej", wierząc, że uwzględnienie wszystkich możliwych wskaźników zmaksymalizuje dokładność.

Takie podejście przypomina popularne błędne przekonanie z zarządzania projektami: "im więcej specjalistów pozyskam, tym bardziej efektywna będzie praca". Jednak, podobnie jak w przypadku zasobów ludzkich, to jakość i narzędzia są ważniejsze niż ilość. Bez uwzględnienia wzajemnych powiązań (korelacji) między danymi lub uczestnikami projektu, zwiększenie objętości może prowadzić jedynie do szumu, zniekształceń, powielania i niepotrzebnego marnotrawstwa.

Ostatecznie często okazuje się, że znacznie bardziej produktywne jest posiadanie mniejszego, ale jakościowo przygotowanego zestawu danych zdolnego do tworzenia stabilnych i rozsądnych prognoz niż poleganie na ogromnych, ale chaotycznych informacjach zawierających wiele sprzecznych sygnałów.

Nadmierna ilość danych nie tylko nie gwarantuje większej dokładności, ale może również prowadzić do zniekształconych wniosków ze względu na obecność szumu, nadmiarowych cech, ukrytych korelacji i nieistotnych informacji. W takich okolicznościach wzrasta ryzyko nadmiernego dopasowania modeli, a wiarygodność wyników analitycznych spada.

W branży budowlanej głównym wyzwaniem związanym z przetwarzaniem dużych zbiorów danych jest określenie optymalnej ilości i jakości danych. Na przykład podczas monitorowania stanu konstrukcji betonowych, korzystanie z tysięcy czujników i zbieranie informacji co minutę może przytłoczyć system przechowywania i analizy danych. Jeśli jednak przeprowadzisz analizę korelacji i wybierzesz 10% najbardziej informatycznych czujników, możesz uzyskać niemal identyczną dokładność prognoz, wydając

wielokrotnie, czasem dziesiątki i setki razy, mniej zasobów.

Korzystanie z mniejszego podzbioru danych zmniejsza zarówno wymaganą ilość pamięci masowej, jak i czas przetwarzania, co znacznie obniża koszty przechowywania i analizy danych i często sprawia, że losowe pobieranie próbek jest idealnym rozwiązaniem do analizy predykcyjnej, szczególnie w przypadku dużych projektów infrastrukturalnych lub pracy w czasie rzeczywistym. Ostatecznie, wydajność procesów budowlanych nie zależy od ilości zebranych danych, ale od jakości ich analizy. Bez krytycznego podejścia i starannej analizy, dane mogą prowadzić do błędnych wniosków.

Po określonej ilości danych każda nowa jednostka informacji przynosi coraz mniej użytecznych wyników. Zamiast bez końca gromadzić informacje, ważne jest, aby skupić się na ich reprezentatywności i metodach analizy (Rys. 9.2-2).

Zjawisko to zostało dobrze opisane przez Allena Wallisa [146], który ilustruje wykorzystanie metod statystycznych na przykładzie testowania dwóch alternatywnych projektów pocisków US Navy.

Marynarka Wojenna przetestowała dwa alternatywne projekty pocisków (A i B), przeprowadzając serię sparowanych rund. W każdej rundzie A otrzymuje 1 lub 0 w zależności od tego, czy jego wydajność jest lepsza czy gorsza od B i odwrotnie. Standardowe podejście statystyczne polega na przeprowadzeniu ustalonej liczby prób (np. 1000) i określeniu zwycięzcy na podstawie rozkładu procentowego (np. jeśli A otrzymuje 1 w ponad 53% przypadków, jest uważany za najlepszy). Kiedy Allen Wallis omówił ten problem z kapitanem (marynarki wojennej) Garrettem L. Schuylerem, kapitan sprzeciwiał się, że taki test, cytując relację Allena, może być bezużyteczny. Gdyby na miejscu był mądry i doświadczony oficer uzbrojenia, taki jak Schuyler, już po pierwszych kilkuset [strzałach] przekonałby się, że eksperymentu nie trzeba przerywać, ani dlatego, że nowa metoda jest wyraźnie gorsza, ani dlatego, że wyraźnie przewyższa to, na co liczono [146].

- Grupa Badań Statystycznych Rządu USA na Uniwersytecie Columbia, okres II wojny światowej

Zasada ta jest szeroko stosowana w różnych branżach. Na przykład w medycynie badania kliniczne nowych leków są przeprowadzane na losowych próbkach pacjentów, co pozwala na uzyskanie statystycznie istotnych wyników bez konieczności testowania leku na całej populacji ludzi żyjących na naszej planecie. W ekonomii i socjologii przeprowadzane są reprezentatywne ankiety w celu odzwierciedlenia opinii społeczeństwa bez konieczności przeprowadzania wywiadów z każdym mieszkańcem kraju.

Podobnie jak rządy i organizacje badawcze badają małe populacje, aby zrozumieć ogólne trendy społeczne, firmy z branży budowlanej mogą wykorzystywać losowe próbki danych do skutecznego monitorowania i tworzenia prognoz do zarządzania projektami (Rys. 9.1-1).

Big data może zmienić podejście do nauk społecznych, ale nie zastąpi statystycznego zdrowego rozsądku [147].

- Thomas Landsall-Welfair, "Forecasting the nation's current mood," *Significance* v. 9(4), 2012 .

Z perspektywy oszczędzania zasobów, podczas gromadzenia danych do przyszłych prognoz i podejmowania decyzji, ważne jest, aby odpowiedzieć na pytanie: czy ma sens wydawanie znaczących zasobów na gromadzenie i przetwarzanie ogromnych zestawów danych, gdy można wykorzystać znacznie mniejszy i tańszy zestaw danych testowych, który można skalować przyrostowo? Skuteczność losowego próbkowania pokazuje, że firmy mogą dziesiątki, a nawet tysiące razy zmniejszyć koszty gromadzenia i szkolenia modeli, wybierając metody gromadzenia danych, które nie wymagają kompleksowego pokrycia, ale nadal zapewniają wystarczającą dokładność i reprezentatywność. Takie podejście pozwala nawet małym firmom osiągać wyniki na równi z dużymi korporacjami przy użyciu znacznie mniejszych zasobów i ilości danych, co jest ważne dla firm, które chcą zoptymalizować koszty i przyspieszyć podejmowanie świadomych decyzji przy użyciu niewielkich zasobów. W kolejnych rozdziałach przedstawiamy przykłady analityki i analityki predykcyjnej opartej na publicznych zbiorach danych z wykorzystaniem narzędzi big data.

Big data: analiza danych ze zbioru milionów pozwoleń na budowę w San Francisco

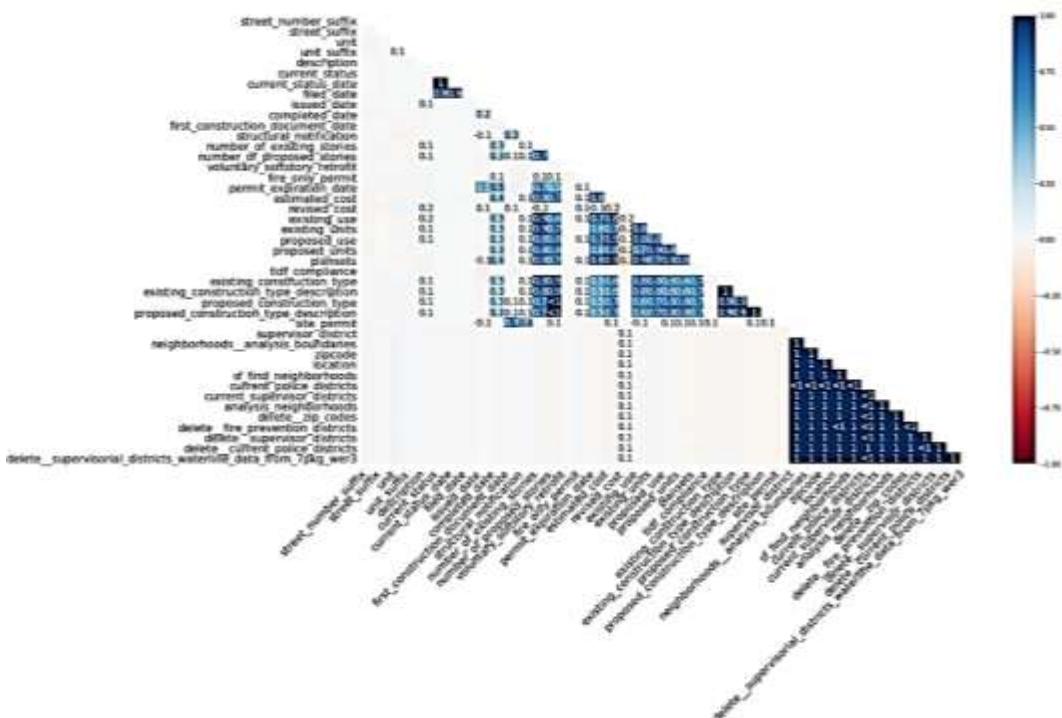
Praca z otwartymi zbiorami danych zapewnia wyjątkową okazję do zastosowania w praktyce zasad omówionych w poprzednich rozdziałach: rozsądnego wyboru cech, reprezentatywnego próbkowania, wizualizacji i krytycznej analizy. W tym rozdziale zbadamy, w jaki sposób złożone zjawiska, takie jak działalność budowlana w dużym mieście, mogą być badane przy użyciu otwartych danych - w szczególności ponad miliona zapisów dotyczących pozwoleń na budowę w San Francisco

Publicznie dostępne dane na temat ponad miliona pozwoleń na budowę (Rys. 9.1-2) (rekordy w dwóch zbiorach danych w formacie CSV) z "San Francisco Department of Buildings" [148] [148] pozwalają nam wykorzystać surową tabelę CSV do analizy nie tylko działalności budowlanej w mieście, ale także do krytycznej analizy najnowszych trendów i historii branży budowlanej San Francisco w ciągu ostatnich 40 lat, od 1980 do 2019 roku.

Przykłady kodu wykorzystane do stworzenia wizualizacji zbioru danych (rysunki 9.1-3- rysunki 9.1-8), a także wykresy wizualne z kodem, objaśnieniami i komentarzami można znaleźć na platformie Kaggle, wyszukując "San Francisco. Construction Sector 1980-2019." [149].

count 1.137695e+06		Building Permits on or after January 1, 2013			Building Permits before January 1, 2013		
permit_creation_date	description	current_status	current_status_date	filed_date	issued_date	completed_date	
07/01/1998	repair stucco	complete		07/07/1998	07/01/1998	07/07/1998	
12/13/2004	reroofing	expired		01/24/2006	12/13/2004	12/13/2004	
02/18/1992	install auto fire spks.	complete		06/29/1992	02/18/1992	03/18/1992	
permit_number	permit_expiration_date	estimated_cost	revised_cost	existing_use	Zipcode	Location	
362780	9812394	11/01/1998	780.0	NaN	1 family dwelling	94123.0 (-37.796246876498, -122.4322541443574)	
570817	200412131233	06/13/2005	9000.0	9000.0	apartments	94127.0 (-37.729258516008388, -122.4644245957462)	
198411	9202396	09/18/1992	9000.0	NaN	apartments	94111.0 (-37.795086002552974, -122.39593224461805)	

Rys. 9.1-2 Zbiory danych zawierają informacje o wydanych pozwoleniach na budowę z różnymi atrybutami obiektów.

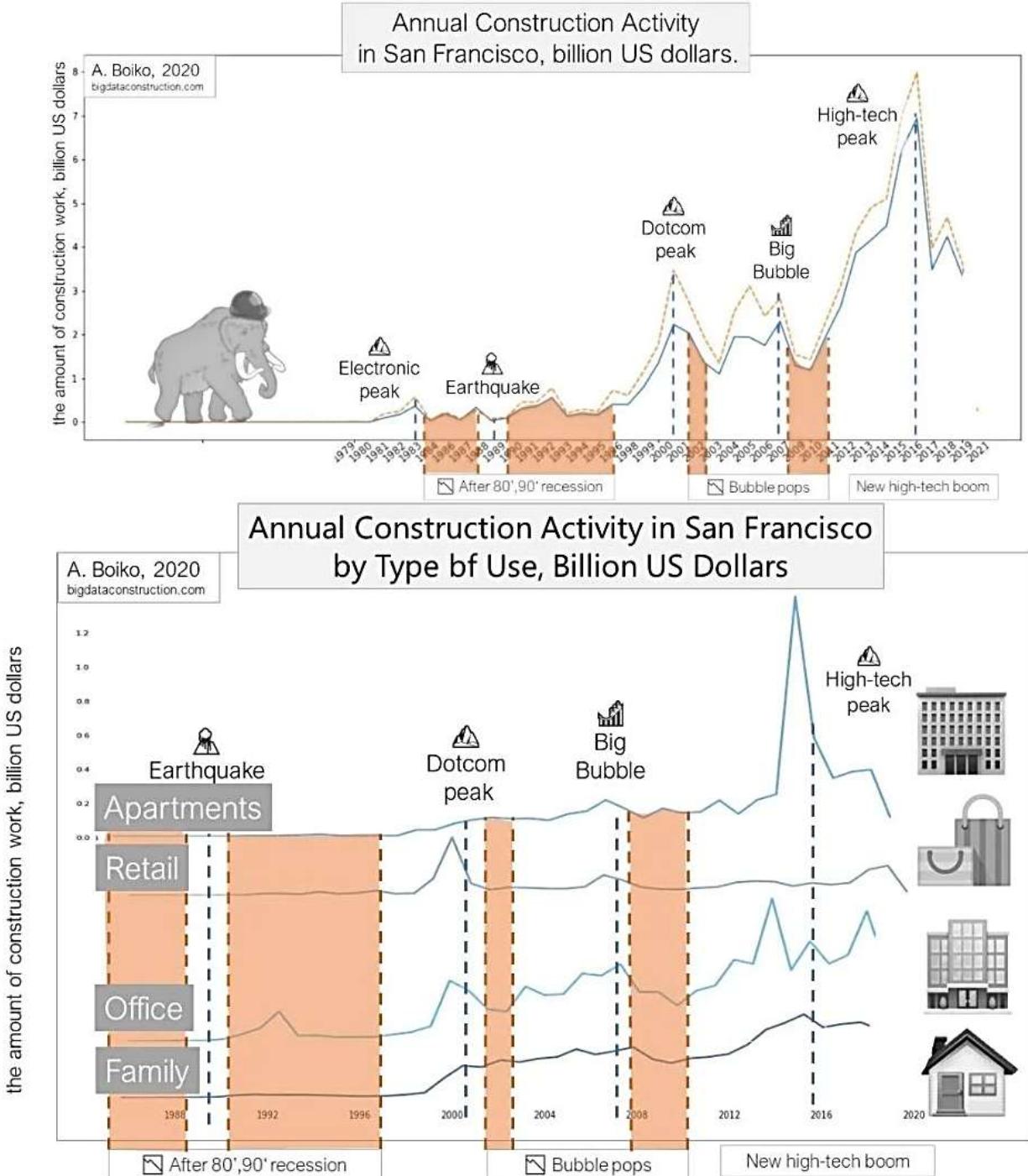


Rys. 9.1-3 Mapa cieplna (Pandas i Seaborn), która wizualizuje wszystkie atrybuty zbioru danych i pomaga zidentyfikować relacje między parami atrybutów.

Z tabeli dostarczonej przez Departament Budynków San Francisco (Rys. 9.1-2) nie wynikają żadne trendy ani wnioski. Suche liczby w formie tabelarycznej nie stanowią podstawy do podejmowania decyzji. Aby dane były zrozumiałe wizualnie, jak omówiono szczegółowo w rozdziałach dotyczących wizualizacji danych, należy je wizualizować za pomocą różnych bibliotek omówionych w części siódmej książki na temat "ETL i wizualizacja wyników w postaci wykresów".

Analizując dane za pomocą Pandas DataFrame i bibliotek wizualizacji Python na podstawie wartości 1 137 695 pozwoleń [148], możemy stwierdzić, że działalność budowlana w San Francisco jest ściśle powiązana z cyklami gospodarczymi, zwłaszcza w dynamicznie rozwijającym się przemyśle technologicznym Doliny Krzemowej (Rys. 9.1-4).

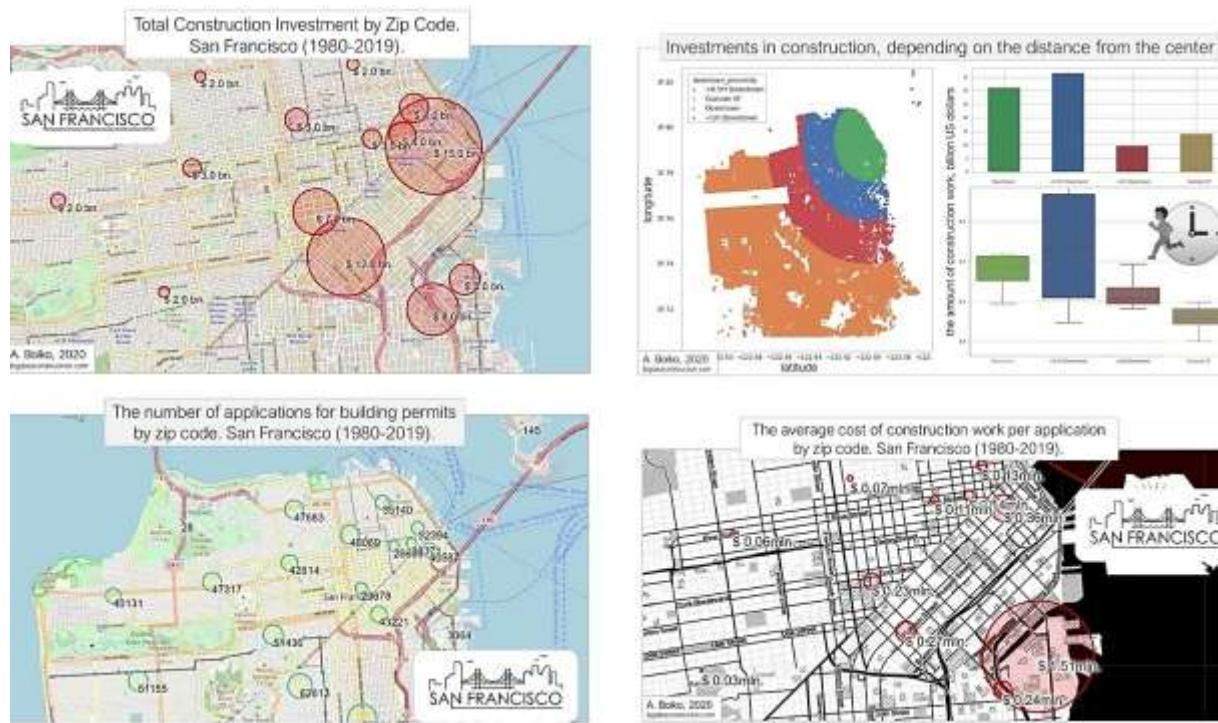
Hossy i bessy gospodarcze mają znaczący wpływ na liczbę i wartość projektów budowlanych. Przykładowo, pierwszy szczyt aktywności budowlanej zbiegły się w czasie z boomem elektronicznym w połowie lat 80-tych (wykorzystano Pandas i Matplotlib), a kolejne szczyty i spadki były związane z bańką internetową i boomem technologicznym ostatnich lat.



Rys. 9.1-4 W sektorze nieruchomości w San Francisco inwestycje są skorelowane z rozwojem technologicznym Doliny Krzemowej.

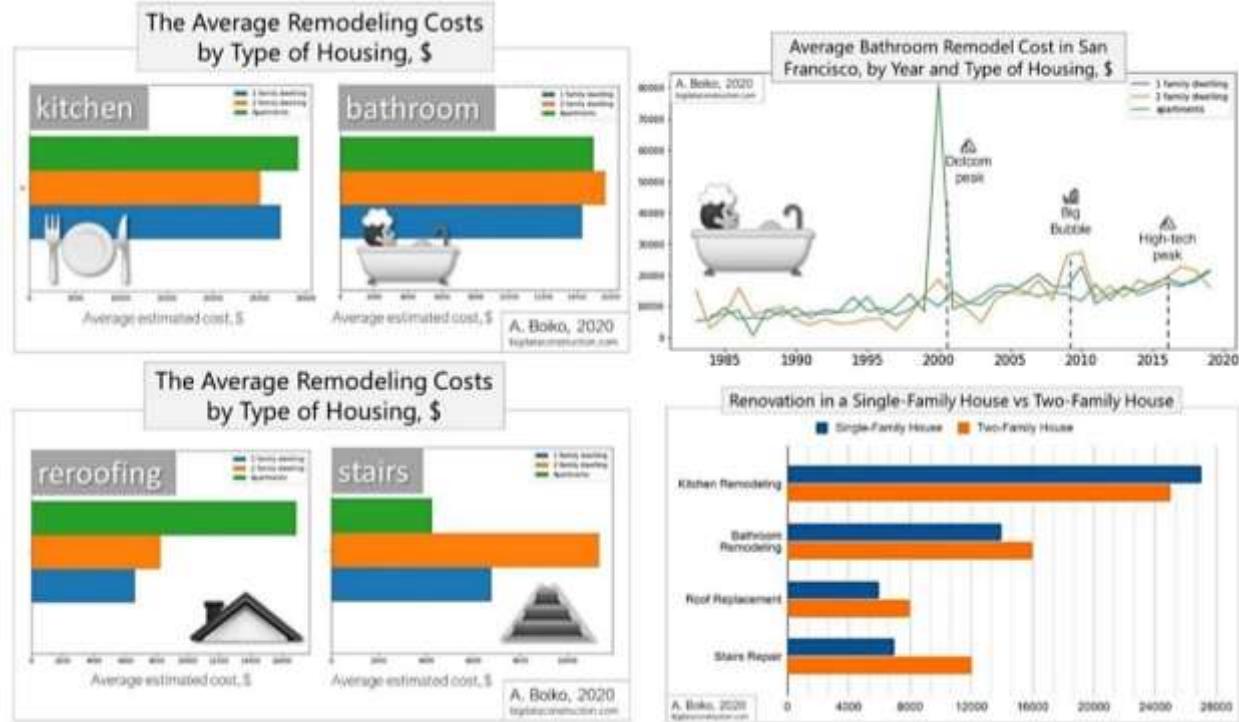
Analizy danych sugerują, że w San Francisco większość z 91,5 mld USD zainwestowanych w budowę i przebudowę w ciągu ostatniej dekady - prawie 75% - koncentruje się w centrum miasta (Rys. 9.1-5 - wykorzystano Pandas i bibliotekę wizualizacji Folium) oraz w promieniu 2 km od centrum miasta, odzwierciedlając większą gęstość inwestycji w tych centralnych strefach.

Średni koszt pozwoleń na budowę różni się znacznie w zależności od dzielnicy, przy czym wnioski w centrum miasta kosztują trzy razy więcej niż te poza centrum miasta ze względu na wyższe koszty gruntów, robocizny, materiałów i surowe przepisy budowlane wymagające droższych materiałów w celu poprawy efektywności energetycznej.



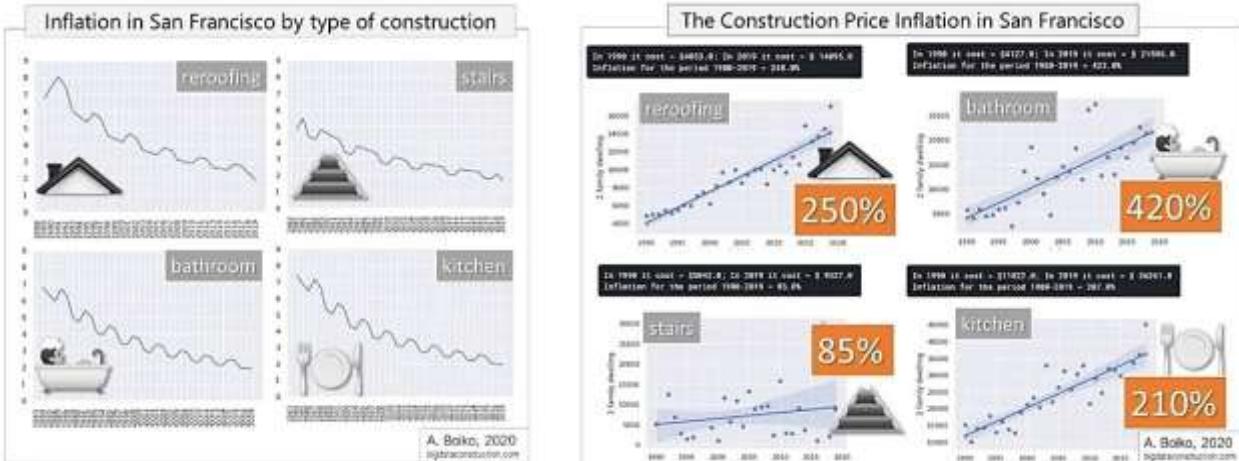
Rys. 9.1-5 W San Francisco 75% inwestycji budowlanych (91,5 mld USD) koncentruje się w centrum miasta.

Zbiór danych pozwala również na obliczenie średnich cen napraw nie tylko według typu domu, ale także według dzielnic miasta i poszczególnych adresów (kodów pocztowych). W San Francisco dynamika kosztów remontów domów wykazuje wyraźne trendy dla różnych rodzajów remontów i mieszkań (Rys. 9.1-6 - wykorzystano Pandas i Matplotlib). Remonty kuchni są wyraźnie droższe niż remonty łazienek: przeciętny remont kuchni w domu jednorodzinnym kosztuje około 28 000 USD w porównaniu do 25 000 USD w domu dwurodzinnym.



Rys. 9.1-6 W SF remonty kuchni kosztują prawie dwa razy więcej niż remonty łazienek, a właściciele domów muszą odkładać \$ 350 każdego miesiąca przez 15 lat, aby pokryć koszty poważnych napraw domu.

Inflację kosztów budowy w San Francisco na przestrzeni lat można prześledzić, analizując dane pogrupowane według typu mieszkania i roku (Rys. 9.1-7 - wykorzystany przez Pandas i Seaborn), który pokazuje stały wzrost średnich kosztów napraw od 1990 roku i ujawnia krótkoterminowe trzyletnie cykle w kosztach napraw budynków wielorodzinnych.

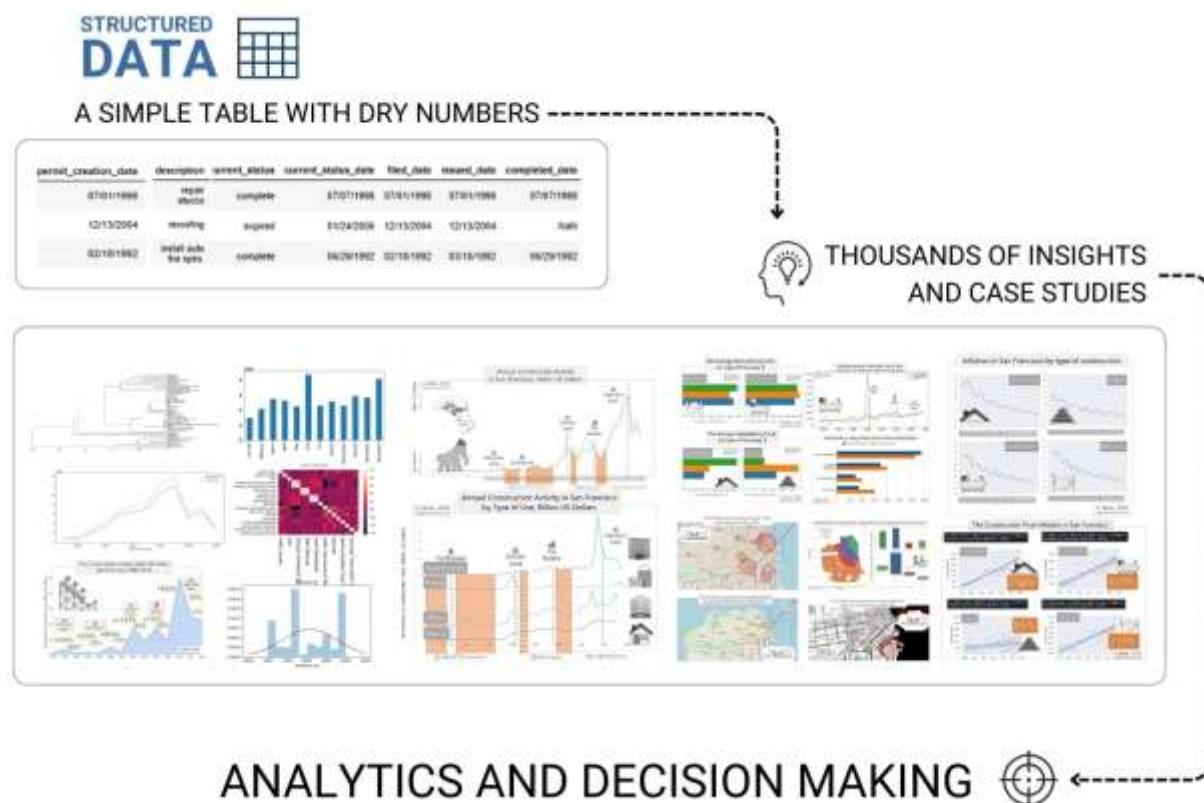


Rys. 9.1-7 Od 1980 r. do 2019 r. koszt remontu łazienki w SF wzrósł pięciokrotnie, podczas gdy cena remontu dachu i kuchni potroiła się, a koszt remontu klatki schodowej wzrósł tylko o 85%.

Badanie publicznych danych z Departamentu Budownictwa San Francisco (Rys. 9.1-3) pokazuje, że koszty

budowy w mieście są niezwykle zmienne i często nieprzewidywalne, na co wpływ ma wiele czynników. Czynniki te obejmują wzrost gospodarczy, innowacje technologiczne i unikalne wymagania różnych typów mieszkań.

W przeszłości takie analizy wymagały dogłębnej znajomości programowania i analityki. Jednak wraz z pojawieniem się narzędzi LLM-, proces ten stał się dostępny i zrozumiał dla szerokiego grona specjalistów w branży budowlanej, od inżynierów w działach projektowych po kierownictwo wyższego szczebla.



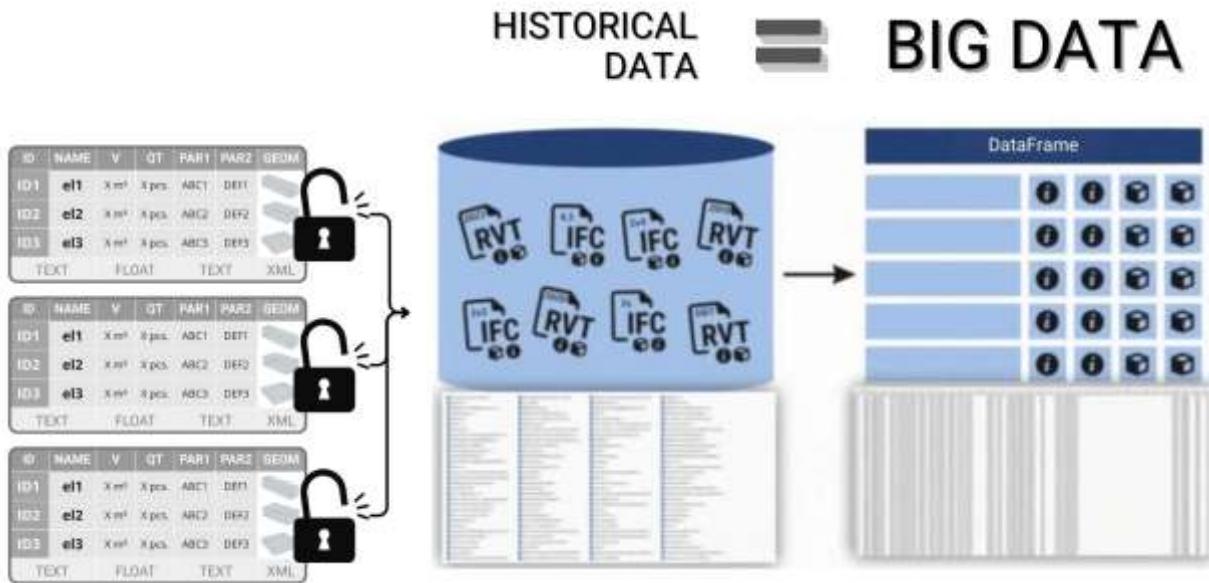
Rys. 9.1-8 Przejście na wizualnie zrozumiałe dane umożliwia zautomatyzowane podejmowanie decyzji poprzez rozpoznawanie ukrytych wzorców.

Tak jak analizowaliśmy dane z tabelarycznego zbioru danych "San Francisco Building Authority", możemy wizualizować i analizować dowolny zbiór danych - od obrazów i dokumentów po dane IoT lub dane z pochodnych baz danych CAD.

Przykład big data w oparciu o dane CAD (BIM)

W poniższym przykładzie przeanalizujemy duży zbiór danych wykorzystujący dane z różnych narzędzi CAD (BIM). Do zebrania i utworzenia dużego zbioru danych wykorzystano wyspecjalizowany automatyczny crawler internetowy (skrypt), skonfigurowany do automatycznego wyszukiwania i zbierania plików projektowych ze stron internetowych oferujących bezpłatne modele architektoniczne w formatach RVT i IFC. W ciągu kilku dni crawler z powodzeniem znalazł i pobrał 4 596 plików IFC, 6 471 plików RVT i 156 024 plików DWG[149].

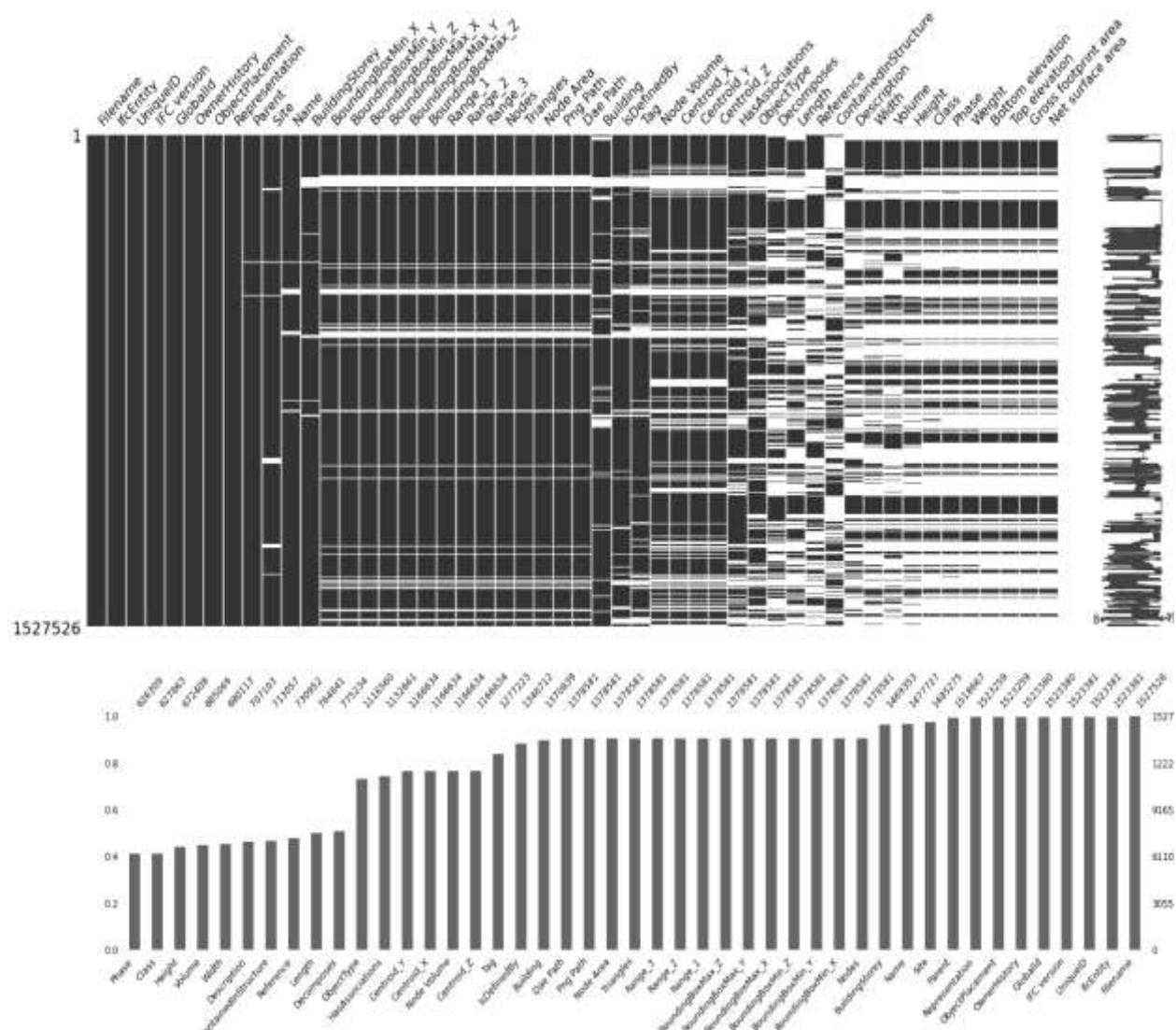
Po zebraniu projektów w formatach RVT i IFC różnych wersji i przekonwertowaniu ich do ustrukturyzowanego formatu CSV przy użyciu bezpłatnych zestawów SDK do inżynierii odwrotnej , prawie 10 tysięcy projektów RVT i IFC zostało zebranych w jeden duży plik tabeli Apache Parquet i przesłanych do Pandas DataFrame w celu analizy (Rys. 9.1-9).



Rys. 9.1-9 Structured Data Dane projektu umożliwiają połączenie dowolnej liczby projektów w jedną dwuwymiarową tabelę.

Dane z tego zbioru na dużą skalę zawierają następujące informacje: zbiór plików IFC zawiera około 4 milionów podmiotów (wierszy) i 24 962 atrybutów (kolumn), a zbiór plików RVT, składający się z około 6 milionów podmiotów (wierszy), zawiera 27 025 różnych atrybutów (kolumn).

Te zbiory informacji (Rys. 9.1-10) obejmują miliony elementów, dla każdego z których dodatkowo uzyskano i dodano do wspólnej tabeli współrzędne geometrii Bounding Box (prostokąt określający granice obiektu w projekcie) oraz obrazy każdego elementu w formacie PNG i geometrii w otwartym formacie XML - DAE (Collada).



Rys. 9.1-10 Subset 1,5 miliona elementów i wizualizacja (biblioteka missingno) zajętości pierwszych 100 atrybutów w postaci histogramu.

W ten sposób uzyskaliśmy wszystkie informacje o dziesiątkach milionów elementów z 4 596 projektów IFC i 6 471 projektów RVT, gdzie wszystkie atrybuty-właściwości wszystkich elementów encji i ich geometrii (Bounding Box) zostały przetłumaczone na ustrukturyzowaną formę pojedynczej tabeli (DataFrame) (rys. 9.1-10 - dane na populacjach ramek danych pojawiają się jako histogramy).

Histogramy (rys. 9.1-10, rys. 9.2-6, rys. 9.2-7) wykreślone podczas procesu analizy pozwalają na szybką ocenę gęstości danych i częstotliwości występowania wartości w kolumnach. Daje to pierwszy wgląd w rozkład cech, obecność wartości odstających i potencjalną przydatność poszczególnych atrybutów w analizie i budowaniu modeli uczenia maszynowego.

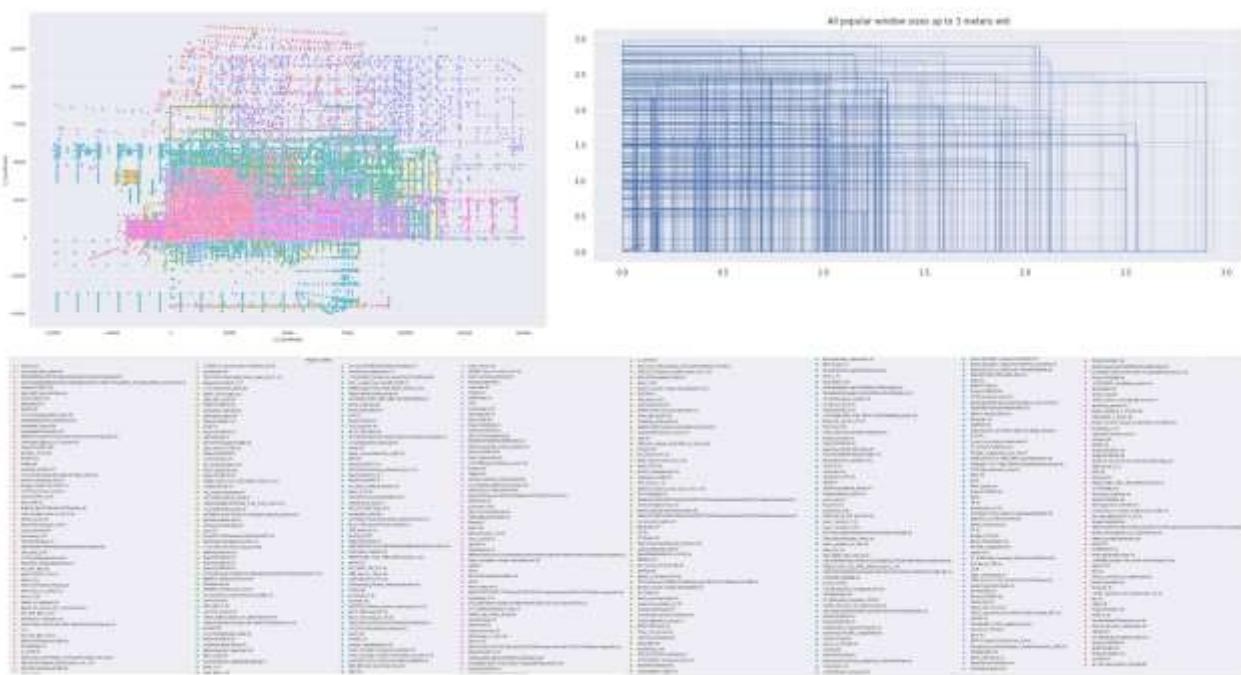
Jednym z przykładów praktycznego wykorzystania tego zbioru danych (rys. 9.1-10) jest projekt "5000 IFC and RVT ". [149], dostępny na platformie Kaggle. Przedstawia on Jupyter Notebook z kompletnym rozwiązaniem Pipeline: od wstępnego przetwarzania i analizy danych po wizualizację wyników przy użyciu

bibliotek Pythona - pandas, matplotlib, seaborn, folium i innych (rys. 9.1-11).



Rys. 9.1-11 Przykłady analizy danych z formatów CAD (BIM) przy użyciu bibliotek wizualizacyjnych Python i biblioteki pandas.

W oparciu o metainformacje możliwe jest określenie, w których miastach opracowano określone projekty i wyświetlenie tego na mapie (np. przy użyciu biblioteki folium). Ponadto znaczniki czasu w danych pozwalają na zbadanie wzorców, kiedy pliki zostały zapisane lub edytowane: według dnia tygodnia, pory dnia i miesiąca.



Rys. 9.1-12 Wizualizacja położenia geometrycznego wszystkich kolumn i wymiarów wszystkich okien do 3 metrów w projektach z listy w dolnej części wykresu.

Parametry geometryczne w postaci Bounding Box wyodrębnione z modeli również nadają się do analiz zbiorczych. Na przykład Rys. 9.1-12 przedstawia dwa wykresy: lewy pokazuje rozkład odległości między

kolumnami dla wszystkich projektów względem punktu zerowego, a prawy pokazuje wymiary wszystkich okien o wysokości do 3 metrów w próbce dziesiątek tysięcy elementów okiennych (po pogrupowaniu całego zbioru danych według parametru "Kategoria" z wartością "OST_Windows", "IfcWindows").

Kod analizy Pipeline dla tego przykładu oraz sam zbiór danych są dostępne na stronie Kaggle pod tytułem "5000 IFC and RVT | DataDrivenConstruction.io" [149]. Gotowy Pipeline wraz z zestawem danych można skopiować i uruchomić za darmo online na Kaggle lub offline w jednym z popularnych IDE: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse z wtyczką PyDev, Thonny, Wing IDE, IntelliJ IDEA z wtyczką Python, JupyterLab lub popularnych narzędziach online Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Analityczne spostrzeżenia uzyskane dzięki przetwarzaniu i badaniu ogromnych ilości ustrukturyzowanych danych będą odgrywać kluczową rolę w procesach decyzyjnych w branży budowlanej.

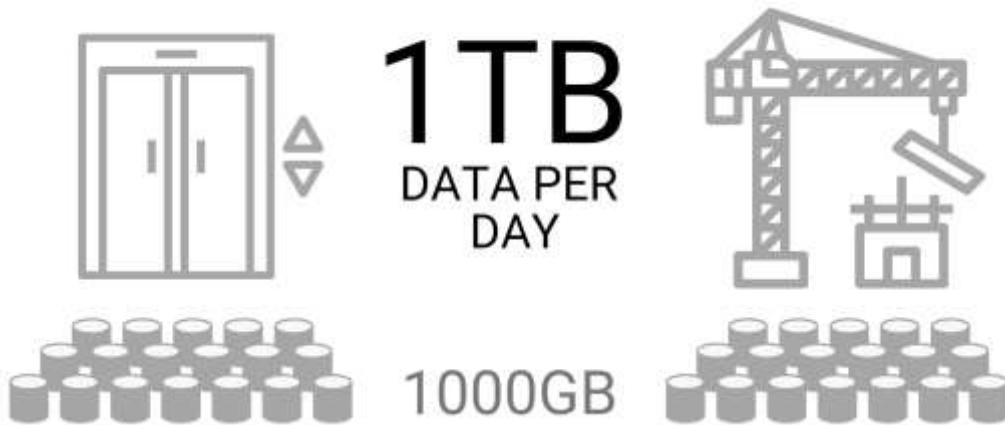
Dzięki tego rodzaju analizie informacji opartej na wcześniejszych projektach, specjalści mogą skutecznie prognozować np. zapotrzebowanie materiały i robociznę oraz optymalizować rozwiązania projektowe przed rozpoczęciem budowy.

Jednakże, podczas gdy dane projektowe lub pozwolenia na budowę są względnie statycznymi informacjami, które zmieniają się stosunkowo wolno, sam proces budowy szybko staje się nasycony różnorodnymi czujnikami i urządzeniami IoT: kamerami, zautomatyzowanymi systemami monitorowania, które przesyłają dane w czasie rzeczywistym - wszystko to zamienia plac budowy w dynamiczne środowisko cyfrowe, w którym dane muszą być analizowane w czasie rzeczywistym.

Internet rzeczy i inteligentne kontrakty IoT

IoT Internet Rzeczy reprezentuje nową falę cyfrowej transformacji, w której każde urządzenie otrzymuje własny adres IP i staje się częścią globalnej sieci. IoT to koncepcja polegająca na połączeniu obiektów fizycznych do Internetu w celu gromadzenia, przetwarzania i przesyłania danych. W budownictwie oznacza to możliwość kontrolowania procesów budowlanych w czasie rzeczywistym, minimalizowania marnotrawstwa materiałów, przewidywania zużycia sprzętu i automatyzacji podejmowania decyzji.

Zgodnie z artykułem CFMA "Przygotowanie do przyszłości dzięki Connected Construction" [150], branża budowlana przejdzie poważną transformację cyfrową w ciągu następnej dekady, której kulminacją będzie koncepcja Connected Construction - w pełni zintegrowanego i zautomatyzowanego placu budowy.



Rys. 9.1-13 IoT lub urządzenia do transmisji danych na placu budowy mogą generować i przesyłać terabajty danych dziennie.

Cyfrowy plac budowy oznacza, że wszystkie elementy budowy - od planowania i logistyki po wykonanie prac i kontrolę jakości na placu budowy przy użyciu kamer stacjonarnych i quadrocopterów - zostaną zintegrowane w jeden dynamiczny cyfrowy ekosystem. Wcześniej, w części 7 tej książki, przyjrzyliśmy się już możliwościom Apache NiFi (Rys. 7.4-5), bezpłatnego i otwartego narzędzia, które umożliwia strumieniowe przesyłanie danych w czasie rzeczywistym - od gromadzenia z różnych źródeł po przesyłanie do platform pamięci masowej lub platform analitycznych.

Dane dotyczące postępu budowy, zużycia materiałów, stanu sprzętu i bezpieczeństwa będą przesyłane w czasie rzeczywistym do systemów analitycznych (rys. 9.1-13). Umożliwia to przewidywanie potencjalnych zagrożeń, szybkie reagowanie na odchylenia i optymalizację procesów na placu budowy. Kluczowe elementy cyfrowego placu budowy obejmują:

- IoT -czujniki - śledzenie parametrów środowiskowych, monitorowanie sprzętu budowlanego i kontrolowanie warunków pracy.
- Cyfrowe bliźniaki - wirtualne modele budynków i infrastruktury w celu przewidywania możliwych odchyleń i zapobiegania błędom.
- Zautomatyzowane systemy logistyczne - zarządzanie łańcuchem dostaw w czasie rzeczywistym w celu ograniczenia przestojów i kosztów.
- Zrobotyzowane systemy budowlane - wykorzystanie autonomicznych maszyn do wykonywania rutynowych i niebezpiecznych zadań.

Robotyzacja wykorzystanie IoT i koncepcja cyfrowego placu budowy Connected Site (Construction) nie tylko zwiększą wydajność i obniżą koszty, ale także zapoczątkują nową erę bezpieczeństwa, zrównoważonego budownictwa i predykcyjnego zarządzania projektami.

Tagi RFID (Radio Frequency Identification) są również jednym z najważniejszych elementów IoT. Są one wykorzystywane do identyfikacji i śledzenia materiałów, maszyn, a nawet personelu na placu budowy, zwiększając przejrzystość i kontrolę zasobów projektu.

Technologia RFID służy do automatycznego rozpoznawania obiektów za pomocą sygnałów radiowych. Składa się ona z trzech kluczowych elementów:

- RFID - znaczniki (pasywne lub aktywne) - zawierają unikalny identyfikator i są przyjmowane do materiałów, narzędzi lub maszyn.
- Skanery to urządzenia, które odczytują informacje z tagów i przesyłają je do systemu.
- Centralizowana baza danych - przechowuje informacje o lokalizacji, statusie i ruchu obiektów.

Zastosowanie RFID w budownictwie:

- Automatyczne rozliczanie materiałów - etykiety na gotowych wyrobach betonowych, zbrojeniach lub opakowaniach płyt warstwowych umożliwiają kontrolę zapasów i zapobiegają kradzieżom.
- Kontrola pracy personelu - RFID - identyfikatory pracowników rejestrują czas rozpoczęcia i zakończenia zmiany, zapewniając rejestr godzin pracy.
- Monitorowanie sprzętu - RFID - system śledzi ruch sprzętu, zapobiegając przestojom i poprawiając wydajność logistyki.

Uzupełnieniem tego pakietu technologii są inteligentne kontrakty oparte na technologii blockchain, które automatyzują płatności, kontrolę dostaw i zgodność z umową bez potrzeby korzystania z pośredników, zmniejszając ryzyko oszustw i opóźnień.

Obecnie, przy braku wspólnego modelu danych, inteligentne kontrakty są po prostu kodem, na który zgadzają się uczestnicy. Jednak dzięki podejściu skoncentrowanemu na danych możliwe jest stworzenie wspólnego modelu parametrów umowy, zakodowanie go w łańcuchu bloków i zautomatyzowanie wypełniania warunków.

Przykładowo, w systemie zarządzania łańcuchem dostaw, inteligentny kontrakt może śledzić dostawę przesyłki na podstawie danych z czujników IoT i tagów RFID oraz automatycznie przekazywać płatność po jej dostarczeniu. Podobnie, na placu budowy, inteligentny kontrakt może rejestrować zakończenie etapu prac - takich jak instalacja prętów zbrojeniowych lub wylewanie fundamentów - w oparciu o dane z dronów lub czujników budowlanych i automatycznie inicjuować kolejną płatność na rzecz wykonawcy bez potrzeby ręcznych kontroli i papierowych certyfikatów.

Jednak pomimo nowych technologii i wysiłków międzynarodowych organizacji normalizacyjnych, mnogość konkurencyjnych ze sobą standardów komplikuje krajobraz IoT.

Według badania Cisco opublikowanego w 2017 roku [151], prawie 60% inicjatyw związanych z Internetem Rzeczy (IoT) zatrzymuje się na etapie weryfikacji koncepcji, a tylko 26% firm uważa swoje projekty IoT za w pełni udane. Co więcej, jedna trzecia ukończonych projektów nie osiąga założonych celów i nie jest uznawana za udane nawet po wdrożeniu.

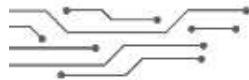
Jednym z kluczowych powodów jest brak interoperacyjności między platformami przetwarzającymi dane z różnych czujników. W rezultacie dane pozostają odizolowane w ramach oddzielnych rozwiązań. Alternatywą dla tego podejścia, podobnie jak w innych podobnych przypadkach (które omówiliśmy w tej książce), jest architektura zbudowana wokół samych danych jako głównego zasobu.

Czujniki IoT odgrywają kluczową rolę nie tylko w monitorowaniu stanu technicznego sprzętu, ale także w analizie predykcyjnej w celu zmniejszenia ryzyka na placu budowy i poprawy ogólnej wydajności procesu poprzez przewidywanie awarii i odchyleń.

Dane gromadzone przez czujniki IoT i znaczniki RFID mogą być przetwarzane w czasie rzeczywistym przez algorytmy uczenia maszynowego, które mogą wykrywać anomalie i z wyprzedzeniem ostrzegać inżynierów o potencjalnych awariach. Może to obejmować zarówno mikropęknięcia w konstrukcjach betonowych, jak i nietypowe przerwy w pracy żurawia wieżowego, wskazujące na awarie techniczne lub naruszenia przepisów. Co więcej, zaawansowane algorytmy analizy behawioralnej mogą wychwytywać wzorce zachowań, które mogą wskazywać na przykład na fizyczne zmęczenie personelu, usprawniając proaktywne zarządzanie bezpieczeństwem i dobrostanem pracowników na miejscu.

W branży budowlanej wypadki i awarie - czy to maszyn, czy ludzi - rzadko zdarzają się nagle. Zwykle poprzedzają je drobne odchylenia, które pozostają niezauważone. Analityka predykcyjna i uczenie maszynowe umożliwiają wykrywanie tych sygnałów na wczesnym etapie, nawet przed wystąpieniem krytycznych konsekwencji.

Podczas gdy dokumenty, pliki projektowe i dane z urządzeń IoT i tagów RFID tworzą cyfrowy ślad projektów budowlanych, uczenie maszynowe może pomóc wydobyć z nich przydatne informacje. Wraz ze wzrostem ilości danych i demokratyzacją dostępu do nich, branża budowlana zyskuje nowe możliwości w zakresie analityki, analityki predykcyjnej i zastosowań sztucznej inteligencji.

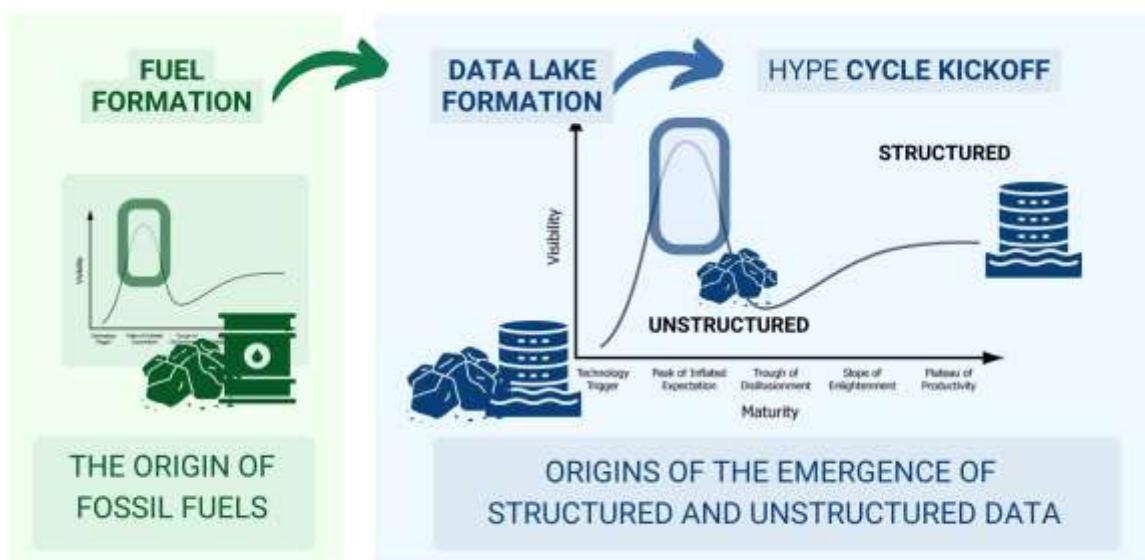


ROZDZIAŁ 8.2.

UCZENIE MASZYNOWE I PROGNOZY

Uczenie maszynowe i sztuczna inteligencja zmienią sposób, w jaki budujemy

Bazy danych różnych systemów w branży budowlanej - z ich nieuchronnie rozpadającą się i coraz bardziej złożoną infrastrukturą - stają się pożywką dla przyszłych rozwiązań. Serwery firmowe, podobnie jak las, są bogate w biomasę ważnych informacji, często ukrytych pod ziemią, w trzewiach folderów i serwerów. Masy danych z różnych systemów tworzonych obecnie - po użyciu, po opadnięciu na dno serwera i po latach skamienienia - będą w przyszłości napędzać uczenie maszynowe i modele językowe. Wewnętrzne czaty firmowe (np. oddzielna instancja lokalnie skonfigurowanego ChatGPT, LLaMa, Mistral, DeepSeek) zostaną zbudowane na tych wewnętrznych modelach przy użyciu scentralizowanej pamięci masowej, aby szybko i wygodnie pobierać informacje i generować niezbędne wykresy, pulpity nawigacyjne i dokumenty.



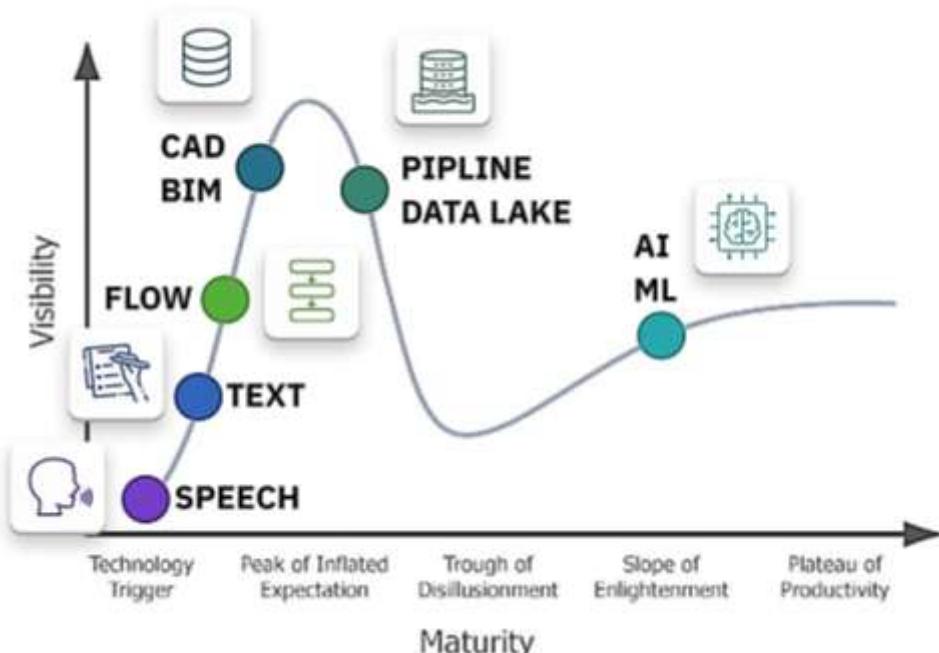
Rys. 9.2-1 Podobnie jak drzewa zamieniają się w węgiel, tak samo informacje z czasem zamieniają się w cenną energię biznesową pod presją czasu i analityki.

Skamienienie masy roślinnej w połączeniu z ciśnieniem i temperaturą tworzy jednorodną i unikalnie ustrukturyzowaną jednorodną masę drzew różnych gatunków, które żyły w różnym czasie - węgiel drzewny [152]. W ten sam sposób informacje zapisane na dyskach twardych w różnych formatach i w różnym czasie pod presją działań analitycznych i temperatury zarządzania jakością ostatecznie tworzą jednorodną ustrukturyzowaną masę cennych informacji (rys. 9.2-1).

Te warstwy (lub częściej pojedyncze bryłki) informacji są tworzone poprzez żmudną organizację danych przez doświadczonych analityków, którzy zaczynają stopniowo wyodrębniać cenne informacje z pozornie długo nieistotnych danych.

W momencie, gdy te dojrzałe warstwy danych nie są już tylko "spalone" w raportach, ale zaczynają krążyć w procesach biznesowych, wzbogacając decyzje i usprawniając procesy, firma staje się gotowa na kolejny krok - przejście do uczenia maszynowego i sztucznej inteligencji (rys. 9.2-2).

Uczenie maszynowe (ML - Machine Learning) to klasa metod rozwiązywania problemów sztucznej inteligencji. Algorytmy uczenia maszynowego rozpoznają wzorce w dużych zbiorach danych i wykorzystują je do samodzielnego uczenia się. Każdy nowy zestaw danych pozwala algorytmom matematycznym na poprawę i dostosowanie się do uzyskanych informacji, co pozwala na ciągłą poprawę dokładności rekomendacji i prognoz.



Rys. 9.2-2 Zanikanie technologii tworzenia danych i stosowanie narzędzi analitycznych otwiera drzwi do tematu uczenia maszynowego.

Jak powiedział wpływowy dyrektor generalny największego na świecie funduszu inwestycyjnego (który posiada kluczowe udziały w prawie wszystkich największych firmach zajmujących się oprogramowaniem budowlanym, a także w firmach, które są właścicielami największej ilości nieruchomości na świecie [55]) w wywiadzie z 2023 r. - uczenie maszynowe zmieni świat budownictwa.

Al ma ogromny potencjał. Zmieni sposób, w jaki pracujemy i żyjemy. Sztuczna inteligencja i robotyka zmienią sposób, w jaki pracujemy i budujemy, a my będziemy mogli wykorzystać sztuczną inteligencję i robotykę jako sposób na osiągnięcie znacznie większej produktywności [153].

- Dyrektor generalny największego na świecie funduszu inwestycyjnego, wywiad, wrzesień 2023 r.

Machine Learning (ML) działa poprzez przetwarzanie dużych ilości danych, wykorzystując techniki statystyczne do naśladowania aspektów ludzkiego myślenia. Jednak większość firm nie posiada takich zbiorów danych, a jeśli tak, to często nie są one wystarczająco oznakowane. W tym miejscu pomocne mogą

być technologie semantyczne i uczenie transferowe, technika, która pozwala ML być bardziej skutecznym w radzeniu sobie z niewielkimi ilościami danych, których wykonalność została omówiona w poprzednich rozdziałach tej części.

Istotą transferu wiedzy jest to, że zamiast przetwarzać każde zadanie od podstaw, można wykorzystać wiedzę zdobytą w pokrewnych dziedzinach. Konieczne jest uświadomienie sobie, że wzorce i odkrycia z innych branż mogą być adaptowane i stosowane w branży budowlanej. Przykładowo, metody optymalizacji procesów logistycznych opracowane w handlu detalicznym pomagają poprawić efektywność zarządzania łańcuchem dostaw w budownictwie. Analiza dużych zbiorów danych, która jest aktywnie wykorzystywana w finansach, może być stosowana do prognozowania kosztów i zarządzania ryzykiem w projektach budowlanych. Rozwijane w przemyśle technologie wizji komputerowej i robotyki już teraz znajdują zastosowanie w zautomatyzowanej kontroli jakości, monitorowaniu bezpieczeństwa i zarządzaniu obiekttami budowlanymi.

Transfer learning pozwala nie tylko przyspieszyć wprowadzanie innowacji, ale także obniżyć koszty ich rozwoju, wykorzystując już zgromadzone doświadczenie innych branż.

$$\text{labor productivity in construction} = f(\text{AI})$$

Rys. 9.2-3 Sztuczna inteligencja technologie i robotyka będą w przyszłości główną siłą napędową zwiększającą produktywność w branży budowlanej.

Ludzkie myślenie jest zorganizowane na podobnej zasadzie: opieramy się na wcześniejszej zdobytej wiedzy, aby rozwiązywać nowe problemy (rys. 4.4-19, rys. 4.4-20, rys. 4.4-21). W uczeniu maszynowym to podejście również się sprawdza - upraszczając model danych i czyniąc go bardziej eleganckim, możemy zmniejszyć złożoność problemu dla algorytmów ML. To z kolei zmniejsza zapotrzebowanie na duże ilości danych i obniża koszty obliczeniowe.

Od subiektywnej oceny do statystycznej prognozy

Era, w której decyzje strategiczne zależały od intuicji poszczególnych menedżerów (Rys. 9.2-4), należy już do przeszłości. W coraz bardziej konkurencyjnym i wymagającym środowisku gospodarczym subiektywne podejście staje się zbyt ryzykowne i nieefektywne. Firmy, które nadal polegają na osobistych opiniach zamiast obiektywnie analizować dane zdolność do szybkiego reagowania na zmiany.

Konkurencyjne środowisko wymaga dokładności i powtarzalności opartej na danych, wzorcach statystycznych i obliczalnym prawdopodobieństwie. Decyzje nie mogą już opierać się na przeczuciach, muszą być oparte na korelacjach, trendach i modelach predykcyjnych wywodzących się z analityki i uczenia maszynowego. To nie tylko zmiana narzędzi - to zmiana logiki myślenia: od założeń do dowodów, od subiektywnych prawdopodobieństw do statystycznie obliczonych odchyleń, od odczuć do faktów.



Rys. 9.2-4 Era decyzji podejmowanych przez HiPPO (opinia najlepiej opłacanego pracownika) odejdzie w przeszłość wraz z pojawiением się dużych zbiorów danych i uczenia maszynowego.

Menedżerowie, którzy wcześniej polegali wyłącznie na własnych odczuciach, nieuchronnie staną w obliczu nowej rzeczywistości: autorytet nie determinuje już wyborów. W centrum zarządzania znajdują się obecnie systemy, które analizują miliony parametrów i wektorów, identyfikując ukryte wzorce i sugerując optymalne strategie.

Głównym powodem, dla którego firmy wciąż unikają wdrażania ML, jest brak przejrzystości. Większość modeli działa jako "czarne skrzynki" dla menedżerów, nie wyjaśniając, w jaki sposób dokładnie dochodzą do swoich wniosków. Prowadzi to do problemów: algorytmy mogą wzmacniać stereotypy, a nawet tworzyć humorystyczne sytuacje, jak w przypadku chatботa Microsoftu, który szybko stał się toksycznym narzędziem komunikacji [154].

W książce Deep Thinking Garry Kasparov, były mistrz świata w szachach, zastanawia się nad swoją porażką z komputerem IBM Big Blue [155]. Argumentuje on, że prawdziwa wartość AI nie polega na kopiowaniu ludzkiej inteligencji, ale na uzupełnianiu naszych umiejętności. Sztuczna inteligencja powinna wykonywać zadania, w których ludzie są słabi, podczas gdy ludzie wnoszą kreatywność. Komputery zmieniły tradycyjne podejście do analizy szachów. Zamiast tworzyć fascynujące historie o grach, komputerowe programy szachowe oceniają każdy ruch bezstronnie, opierając się wyłącznie na jego rzeczywistej sile lub słabości. Kasparow zauważa, że ludzka tendencja do postrzegania wydarzeń jako spójnych historii, a nie pojedynczych działań, często prowadzi do błędnych wniosków - nie tylko w szachach, ale w życiu w ogóle.

Dlatego też, jeśli planujesz wykorzystać uczenie maszynowe do przewidywania i analizy, ważne jest, aby zrozumieć jego podstawowe zasady - jak działają algorytmy i jak przetwarzane są dane - zanim zaczniesz korzystać z narzędzi uczenia maszynowego i AI w swojej pracy. Najlepszym sposobem na rozpoczęcie pracy jest zdobycie praktycznego doświadczenia.

Jednym z najwygodniejszych narzędzi do wstępnego wprowadzenia do tematu uczenia maszynowego i predykcji jest Jupyter Notebook i popularny klasyczny zestaw danych Titanic, który zapewni wizualne wprowadzenie do kluczowych metod analizy danych i budowania modeli ML.

Zbiór danych Titanic: Witaj świecie w świecie analityki dane i duże zbioru danych

Jednym z najbardziej znanych przykładów wykorzystania ML w analityce danych jest analiza zbioru danych Titanica, który jest często wykorzystywany do badania prawdopodobieństwa przeżycia pasażerów. Nauka tej tabeli jest analogiczna do programu "Hello World" podczas nauki języków programowania.

Zatonięcie statku RMS Titanic w 1912 roku spowodowało śmierć 1502 z 2224 osób. Zbiór danych Titanica zawiera nie tylko informacje o tym, czy pasażer przeżył, ale także atrybuty, takie jak: wiek, płeć, klasa biletu i inne parametry. Ten zbiór danych jest dostępny za darmo i może być otwierany i analizowany na różnych platformach offline i online.

Link do zbioru danych Titanica:

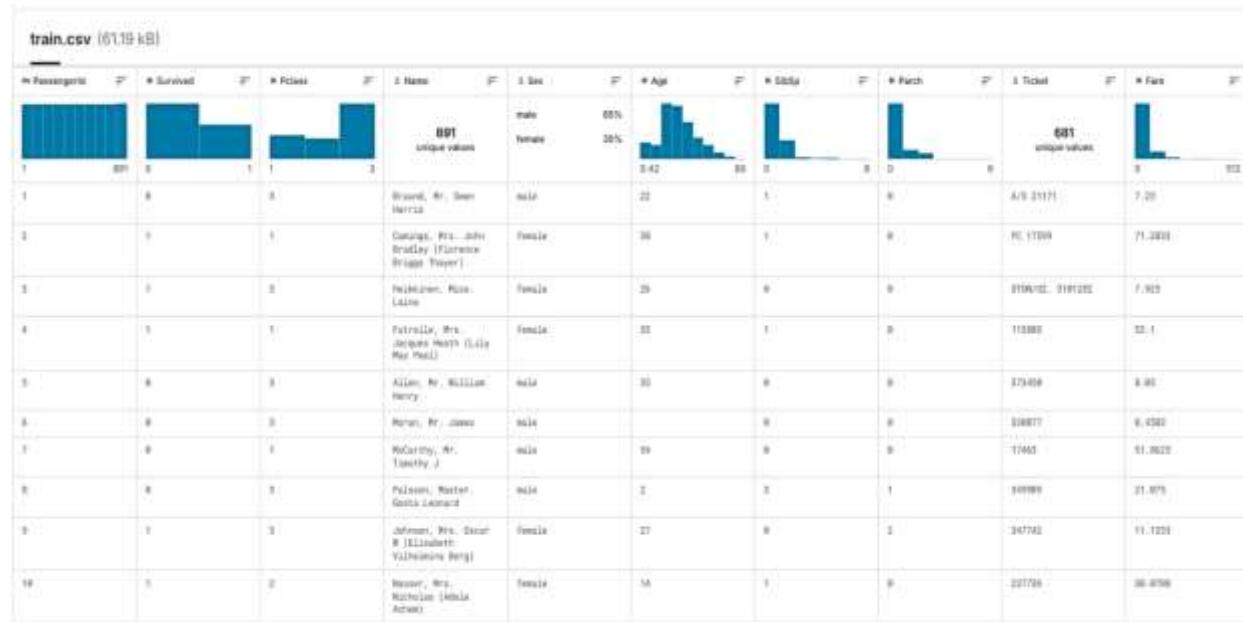
<https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv>

Wcześniej w rozdziale "IDE obsługujące LLM i przyszłe zmiany w programowaniu" omówiliśmy już Jupyter Notebook - jedno z najpopularniejszych środowisk programistycznych do analizy danych i uczenia maszynowego. Bezpłatnymi chmurowymi odpowiednikami Jupyter Notebook są platformy Kaggle i Google Collab, które umożliwiają uruchamianie kodu Python bez instalowania oprogramowania i zapewniają bezpłatny dostęp do zasobów obliczeniowych.

Kaggle to największa platforma do analizy danych i uczenia maszynowego ze zintegrowanym środowiskiem wykonywania kodu. Od października 2023 r. Kaggle ma ponad 15 milionów użytkowników [156] ze 194 krajów.

Pobierz i korzystaj ze zbioru danych Titanic na platformie Kaggle (Rys. 9.2-5), aby przechowywać zbiór

danych (jego kopię) i uruchamiać kod Python z preinstalowanymi bibliotekami bezpośrednio w przeglądarce, bez konieczności instalowania dedykowanego IDE.



Rys. 9.2-5 Statystyki tabeli Titanic - najpopularniejszy treningowy zbiór danych do nauki analityki danych i uczenia maszynowego.

Zbiór danych Titanic zawiera dane dotyczące 2224 pasażerów na pokładzie *RMS Titanic* w czasie jego katastrofy w 1912 roku. Zbiór danych jest przedstawiony jako dwie oddzielne tabele, próbka szkoleniowa (*train.csv*) i próbka testowa (*test.csv*), co pozwala na wykorzystanie go zarówno do szkolenia modeli, jak i do oceny ich dokładności na nowych danych.

Treningowy zbiór danych zawiera zarówno atrybuty - atrybuty pasażerów (wiek, płeć, klasa biletu i inne), jak i informacje o tym, kto przeżył (kolumna z wartościami binarnymi "Przeżył"). Treningowy zbiór danych (Rys. 9.2-6 - plik *train.csv*) jest używany do trenowania modelu. Testowy zbiór danych (rys. 9.2-7 - plik *test.csv*) zawiera tylko atrybuty pasażerów bez informacji o ocalałych (bez pojedynczej kolumny "Survivor"). Testowy zbiór danych ma na celu przetestowanie modelu na nowych danych i ocenę jego dokładności.

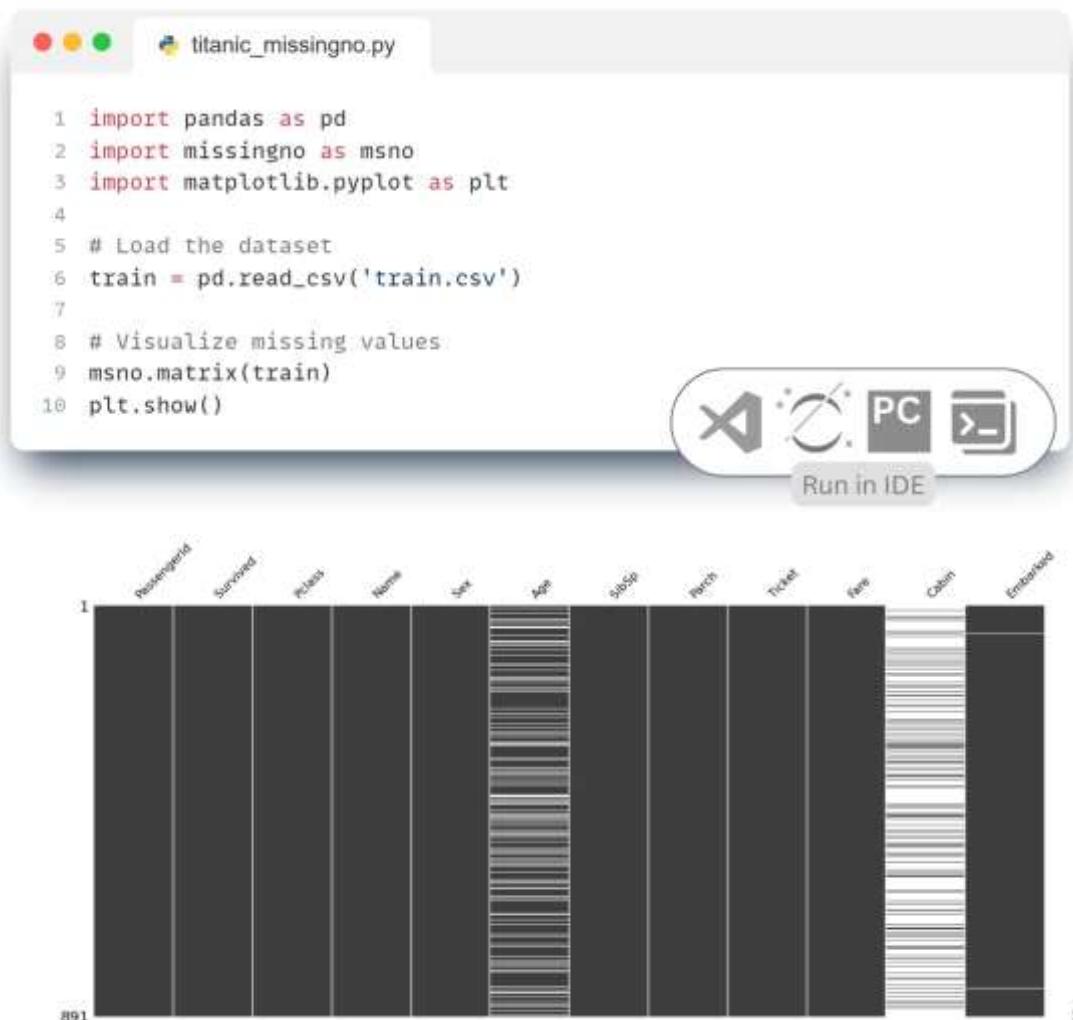
W ten sposób mamy prawie identyczne atrybuty pasażerów w treningowych i testowych zbiorach danych. Jedyną kluczową różnicą jest to, że w testowym zbiorze danych mamy listę pasażerów, którzy nie mają kolumny "Survivor" - zmiennej docelowej, którą chcemy nauczyć się przewidywać za pomocą różnych algorytmów matematycznych. Po zbudowaniu modelu będziemy mogli porównać wynik naszego modelu z rzeczywistym parametrem "Survivor" z testowego zbioru danych, który weźmiemy pod uwagę przy ocenie wyników.

Główne kolumny tabeli, parametry pasażerów w zbiorze danych treningowych i testowych:

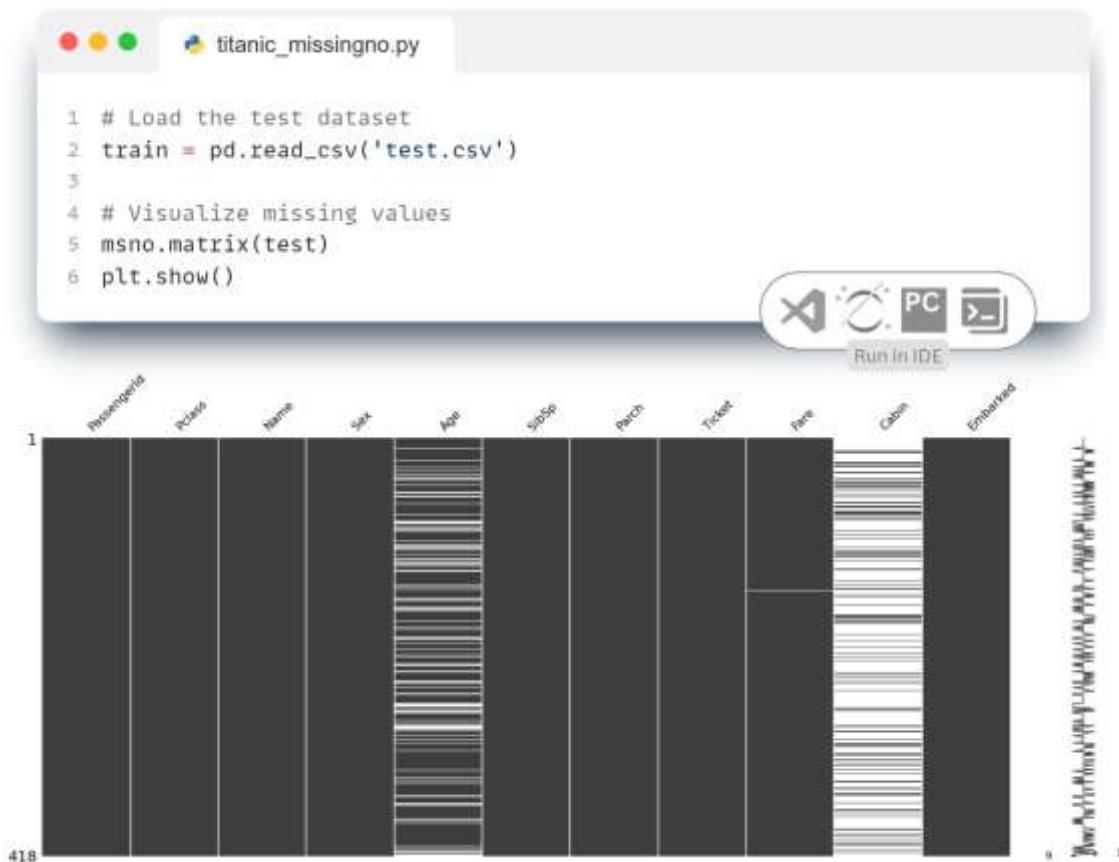
- **PassengerId** - unikalny identyfikator pasażera
- **Przeżył** - 1, jeśli pasażer przeżył, 0, jeśli nie żyje (niedostępne w zestawie testowym).
- **Pclass** - klasa biletu (1, 2 lub 3)
- **Imię i nazwisko** - imię i nazwisko pasażera

- **Płeć** - płeć pasażera (mężczyzna/kobieta)
- Wiek
- **SibSp** - liczba braci/siostr lub małżonków na pokładzie
- **Parch** - liczba rodziców lub dzieci na pokładzie
- **Bilet** - numer biletu
- **Taryfa** - cena biletu
- Kabina - numer kabiny (brakuje wielu danych)
- **Embarked** oznacza port wejścia na pokład (C = Cherbourg, Q = Queenstown, S = Southampton).

Do wizualizacji brakujących danych w obu tabelach można użyć biblioteki missingno (rys. 9.2-6, rys. 9.2-7), która wyświetla brakujące wartości w formie histogramu, gdzie białe pola wskazują brakujące dane. Ta wizualizacja umożliwia szybką ocenę jakości danych przed ich przetworzeniem.



Rys. 9.2-6 Kilka linijek kodu służy do wizualizacji brakujących danych w szkoleniowym zbiorze danych Titanic, gdzie kluczowym parametrem szkolenia jest parametr "Survived".



Rys. 9.2-7 Wizualizacja brakujących danych w testowym zbiorze danych Titanic, który zawiera tylko charakterystykę pasażerów bez żadnych informacji.

Przed sformułowaniem hipotez i dokonaniem prognoz na podstawie zbioru danych, analizy wizualne pomagają zidentyfikować kluczowe wzorce w danych, ocenić ich jakość i zidentyfikować możliwe zależności. Istnieje wiele technik wizualizacji, które mogą pomóc w lepszym zrozumieniu zbioru danych Titanica. Możesz użyć wykresów dystrybucji do analizy grup wiekowych pasażerów, wykresów przeżycia według płci i klasy oraz macierzy brakujących danych, aby ocenić jakość informacji i zrozumieć dane.

- Poprośmy LLM o pomoc w wizualizacji danych ze zbioru Titanic, wysyłając następujące żądanie tekstowe do dowolnego modelu LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN lub dowolnego innego):

Pokaż kilka prostych wykresów dla zbioru danych Titanic. Pobierz zestaw danych samodzielnie i pokaż ↵

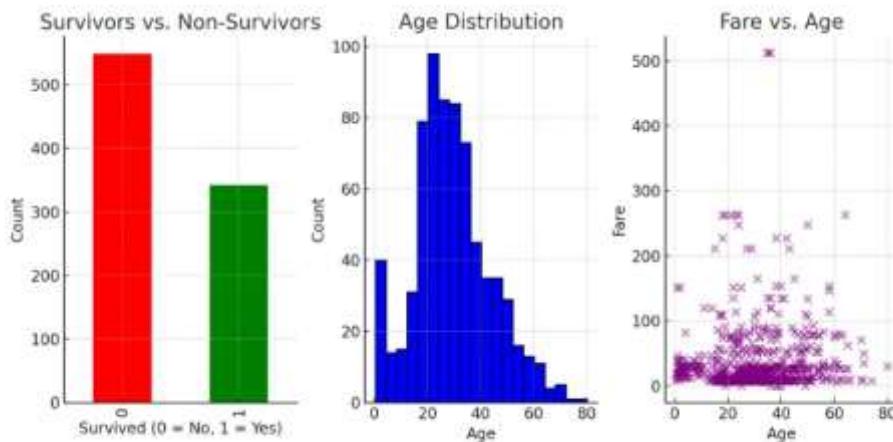
- Odpowiedź LLM w postaci gotowego kodu i wykresów wizualizujących parametry zbioru danych

```

 1 import pandas as pd
 2 import matplotlib.pyplot as plt
 3
 4 # Load the dataset
 5 file_path = "/mnt/data/titanic.csv"
 6 df = pd.read_csv(file_path)
 7
 8 # Convert column names to lowercase for consistency
 9 df.columns = df.columns.str.lower()
10
11 # Basic Plots
12 plt.figure(figsize=(10, 5))
13
14 # Bar Chart: Survival Count
15 plt.subplot(1, 3, 1)
16 df['survived'].value_counts().plot(kind='bar', color=['red', 'green'])
17 plt.title("Survivors vs. Non-Survivors")
18 plt.xlabel("Survived (0 = No, 1 = Yes)")
19 plt.ylabel("Count")
20
21 # Histogram: Age Distribution
22 plt.subplot(1, 3, 2)
23 df['age'].dropna().plot(kind='hist', bins=20, color='blue', edgecolor='black')
24 plt.title("Age Distribution")
25 plt.xlabel("Age")
26 plt.ylabel("Count")
27
28 # Scatter Plot: Fare vs. Age
29 plt.subplot(1, 3, 3)
30 plt.scatter(df['age'], df['fare'], alpha=0.5, color='purple')
31 plt.title("Fare vs. Age")
32 plt.xlabel("Age")
33 plt.ylabel("Fare")
34
35 # Show the plots
36 plt.tight_layout()
37 plt.show()

```

danych



Rys. 9.2-8 LLM pomaga uzyskać natychmiastową wizualizację danych zbioru danych.

Wizualizacja danych jest ważnym krokiem w przygotowaniu zbioru danych do późniejszej budowy modelu uczenia maszynowego, do którego można uzyskać dostęp tylko poprzez zrozumienie danych.

Uczenie maszynowe w akcji: od pasażerów Titanica do zarządzania projektami

Główną hipotezą wykorzystaną do zbadania struktury uczenia maszynowego w oparciu o zbiór danych Titanica jest to, że niektóre grupy pasażerów miały większe szanse na przeżycie.

Mała tabela pasażerów Titanica stała się popularna na całym świecie, a miliony ludzi używają jej do szkolenia, eksperymentowania i testowania modeli, aby dowiedzieć się, które algorytmy i hipotezy zbudują najdokładniejszy model przewidywania przeżycia w oparciu o zestaw danych szkoleniowych dla pasażerów Titanica.

Atrakcyjność zbioru danych Titanic wynika z jego zwartości: z kilkuset wierszami i dwunastoma kolumnami (rys. 9.2-6), zapewnia on szerokie możliwości analizy. Zbiór danych jest klasycznym przykładem rozwiązania klasyfikacji binarnej, gdzie cel problemu - przetrwanie - jest wyrażony w wygodnym formacie 0 lub 1.

John Wheeler w "It from Bit" [7] twierdzi, że wszechświat opiera się na wyborach binarnych. Podobnie, biznes prowadzony przez ludzi składających się z cząsteczek jest w rzeczywistości zbudowany na serii binarnych wyborów binarnych.

Ponadto dane te są oparte na prawdziwym wydarzeniu historycznym, co czyni je cennymi dla badań, w przeciwieństwie do sztucznie stworzonych przykładów. Na samej platformie Kaggle, jednej z największych platform Data Pipeline i ETL, 1 355 998 osób wzięło udział w wyzwaniach opartych na zbiorze danych Titanic, opracowując 53 963 unikalnych rozwiązań Data Pipeline [157] (Rys. 9.2-9).

Wydaje się to niewiarygodne, ale zaledwie 1000 wierszy danych o pasażerach Titanica z 12 parametrami stało się polem dla milionów hipotez, łańcuchów logicznych i unikalnych Data-Pipelines. Z małego zbioru danych rodzą się nieskończone spostrzeżenia, hipotezy i interpretacje - od prostych modeli przetrwania po złożone zespoły, które uwzględniają ukryte wzorce i złożone labirynty rozumowania.

Machine Learning from Disaster

Submit Prediction

Data Code Models Discussion Leaderboard Rules

Titanic Tutorial
Updated 3y ago
29858 comments · Titanic - Machine Learning from Disaster

Titanic competition w/ TensorFlow Decision Forests
Updated 2y ago
Score: 0.80143 · 318 comments · Titanic - Machine Learning from Disaster

Titanic Data Science Solutions
Updated 6y ago
2590 comments · Titanic - Machine Learning from Disaster

Exploring Survival on the Titanic
Updated 7y ago
Score: 0.80382 · 1072 comments · Titanic - Machine Learning from Disaster

Rys. 9.2-9 Pierwsze pięć rozwiązań spośród 53 963 gotowych i otwartych rozwiązań Pipeline. Prawie 1,5 miliona osób próbowało już rozwiązać ten problem tylko na Kaggle [157].

Jeśli nawet tak mała tabela może wygenerować miliony unikalnych rozwiązań (rys. 9.2-9), co możemy powiedzieć o rzeczywistych zestawach danych dotyczących konstrukcji przemysłowych, w których parametry są mierzone w dziesiątkach tysięcy?

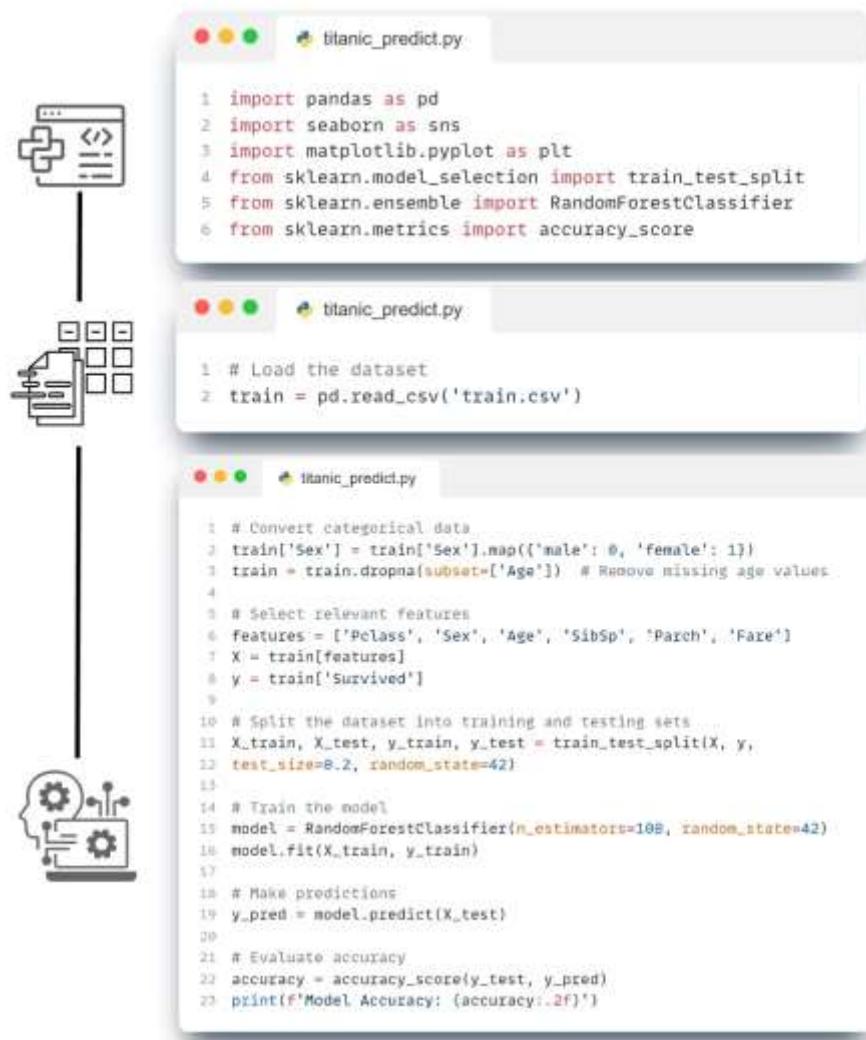
Standardowy projekt CAD stosunkowo niewielkiego budynku zawiera dziesiątki tysięcy elementów o tysiącach parametrów - od cech geometrycznych po koszty i czas. Wyobraź sobie, ile potencjalnych spostrzeżeń, relacji, przewidywań i hipotez dotyczących zarządzania kryje się w danych ze wszystkich projektów Twojej firmy zebranych na przestrzeni lat. Historyczne dane projektowe to nie tylko archiwum - to żywa pamięć organizacji, jej cyfrowy ślad, który można analizować w celu zbudowania dużej liczby unikalnych hipotez.

Co najważniejsze, nie musisz czekać, aż społeczność Kaggle zainteresuje się Twoją firmą lub Twoimi danymi. Możesz zacząć pracować z tym, co masz już dziś: uruchamiać analizy na własnych danych, trenować modele na własnych danych, identyfikować powtórzenia, anomalie i wzorce. Tam, gdzie kiedyś potrzebne były lata eksperymentów i kosztownych konsultacji, teraz wystarczy inicjatywa, LLM, otwarte podejście do danych i chęć uczenia się.

- ➲ Aby zbudować algorytm uczenia maszynowego, który będzie przewidywał przeżywalność pasażerów w oparciu o treningowy zbiór danych pasażerów train.csv, poprośmy LLM o rozwiązanie tego problemu za nas:

W oparciu o zbiór danych treningowych pasażerów Titanica, zbuduj model uczenia maszynowego, aby przewidzieć przeżywalność. ↴

💡 Odpowiedź LLM :



```

titanic_predict.py
1 import pandas as pd
2 import seaborn as sns
3 import matplotlib.pyplot as plt
4 from sklearn.model_selection import train_test_split
5 from sklearn.ensemble import RandomForestClassifier
6 from sklearn.metrics import accuracy_score

titanic_predict.py
1 # Load the dataset
2 train = pd.read_csv('train.csv')

titanic_predict.py
1 # Convert categorical data
2 train['Sex'] = train['Sex'].map({'male': 0, 'female': 1})
3 train = train.dropna(subset=['Age']) # Remove missing age values
4
5 # Select relevant features
6 features = ['Pclass', 'Sex', 'Age', 'SibSp', 'Parch', 'Fare']
7 X = train[features]
8 y = train['Survived']
9
10 # Split the dataset into training and testing sets
11 X_train, X_test, y_train, y_test = train_test_split(X, y,
12 test_size=0.2, random_state=42)
13
14 # Train the model
15 model = RandomForestClassifier(n_estimators=100, random_state=42)
16 model.fit(X_train, y_train)
17
18 # Make predictions
19 y_pred = model.predict(X_test)
20
21 # Evaluate accuracy
22 accuracy = accuracy_score(y_test, y_pred)
23 print(f'Model Accuracy: {accuracy:.2f}')

```

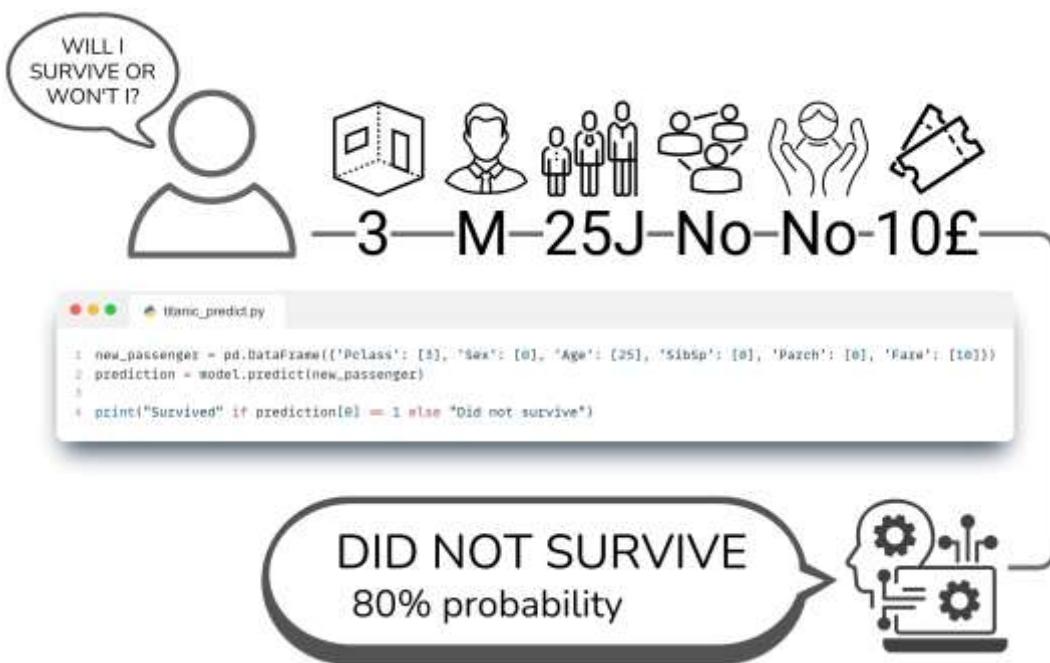
Rys. 9.2-10 LLM zbudował prognozę ocalałych z Titanica przy użyciu algorytmu uczenia maszynowego Random Forest.

Kod wynikowy z LLM (rys. 9.2-10) ładuje dane pasażerów Titanica, czyści je, konwertuje zmienne kategoryczne (np. płeć do formatu liczbowego) i trenuje model za pomocą algorytmu RandomForestClassifier, aby przewidzieć, czy pasażer przeżył, czy nie (więcej o popularnych algorytmach powiemy w kolejnych rozdziałach).

Kod dzieli dane treningowe na zestawy treningowe i testowe (strona internetowa Kaggle utworzyła już gotowe pliki test.csv (rys. 9.2-7) i train.csv (rys. 9.2-6) do treningu, a następnie model jest trenowany na danych treningowych i testowany na danych testowych, aby sprawdzić, jak dobry jest dany model predykcyjny. Po treningu dane testowe z pliku test.csv (z prawdziwymi danymi o osobach, które przeżyły lub nie przeżyły) są wprowadzane do modelu i przewidują, kto przeżył, a kto nie. W naszym przypadku dokładność modelu uczenia maszynowego, który uzyskaliśmy wynosi około 80%, co pokazuje, że dość dobrze oddaje on wzorce.

Uczenie maszynowe można porównać do dziecka próbującego dopasować prostokątny klocek do okrągłego otworu. Na początkowych etapach algorytm próbuje wielu podejść, napotykając błędy i niespójności. Proces ten może wydawać się nieefektywny, ale zapewnia ważną naukę: analizując każdy błąd, model poprawia swoje przewidywania i podejmuje coraz dokładniejsze decyzje.

Teraz ten model (rys. 9.2-10) może być używany do przewidywania wskaźnika przeżycia nowych pasażerów i na przykład, jeśli podasz mu informacje o pasażerach za pomocą funkcji model.predict parametrami: "mężczyzna", "3 klasa", "25 lat", "brak krewnych na pokładzie", model wygeneruje prognozę - że pasażer z 80% prawdopodobieństwem nie przeżyje katastrofy, jeśli był na Titanice w 1912 roku (rys. 9.2-11).



Rys. 9.2-11 Model, który stworzyliśmy powyżej, może teraz przewidzieć z 80% prawdopodobieństwem, czy jakikolwiek nowy pasażer Titanica przeżyje.

Model przewidywania przeżycia pasażerów Titanica ilustruje znacznie szerszą koncepcję: każdego dnia tysiące profesjonalistów w branży budowlanej podejmuje podobne "podwójne" decyzje - o życiu lub śmierci decyzji, projektu, oszacowania, narzędzia, zysku lub straty, bezpieczeństwa lub ryzyka. Podobnie jak w przypadku Titanica, gdzie wynik zależała od czynników (płeć, wiek, klasa), w budownictwie na każdy aspekt decyzji wpływa wiele własnych czynników i zmiennych (kolumny tabel): koszt materiałów, kwalifikacje

pracowników, terminy, pogoda, logistyka, ryzyko techniczne, komentarze i setki tysięcy innych parametrów.

W branży budowlanej uczenie maszynowe opiera się na tych samych zasadach, co w innych dziedzinach: modele są szkolone na danych historycznych - z projektów, kontraktów, szacunków - w celu przetestowania różnych hipotez i znalezienia najbardziej efektywnych rozwiązań. Proces ten przypomina uczenie dziecka metodą prób i błędów: z każdym cyklem modele dostosowują się i stają się coraz dokładniejsze.

Wykorzystanie zgromadzonych danych otwiera nowe horyzonty dla budownictwa. Zamiast czasochłonnych ręcznych obliczeń, można trenować modele, które mogą przewidywać kluczowe cechy przyszłych projektów z dużą dokładnością. W ten sposób analityka predykcyjna przekształca branżę budowlaną w przestrzeń, w której można nie tylko planować, ale także pewnie przewidywać rozwój wydarzeń.

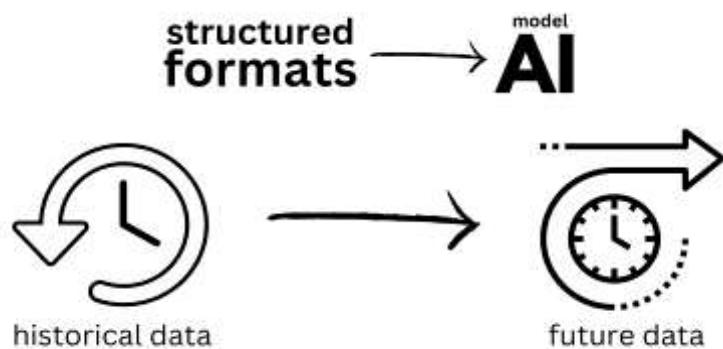
Przewidywania i prognozy oparte na danych historycznych

Dane zebrane na temat projektów firmy otwierają możliwość budowania modeli zdolnych do przewidywania kosztów i charakterystyki czasowej przyszłych, jeszcze niezrealizowanych obiektów - bez czasochłonnych ręcznych obliczeń i porównań. Pozwala to znacznie przyspieszyć i uprościć procesy wyceny, opierając się nie na subiektywnych założeniach, ale na solidnych prognozach matematycznych.

Wcześniej, w czwartej części książki, rozważaliśmy szczegółowo tradycyjne metody szacowania kosztów projektu, w tym metodę opartą na zasobach, a także wspomnieliśmy o podejściach parametrycznych i eksperckich. Metody te są nadal istotne, ale w nowoczesnej praktyce zaczynają być wzbogacane o narzędzia analizy statystycznej i uczenia maszynowego, które mogą znacznie poprawić dokładność i powtarzalność szacunków.

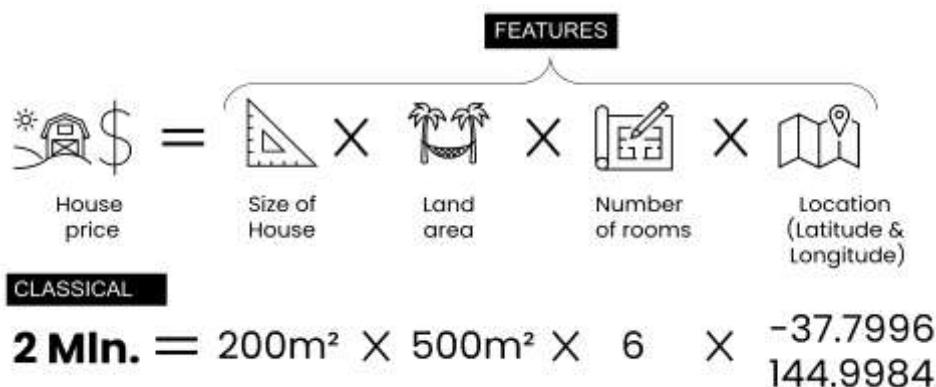
Procesy ręcznego i półautomatycznego obliczania cen i atrybutów czasowych zostaną w przyszłości uzupełnione o opinie i prognozy modeli ML zdolnych do analizowania danych historycznych, znajdowania ukrytych wzorców i proponowania świadomych decyzji. Nowe dane i scenariusze będą generowane automatycznie na podstawie już dostępnych informacji - podobnie jak modele językowe (LLM) tworzą teksty, obrazy i kod na podstawie danych gromadzonych przez lata z otwartych źródeł [158].

Tak jak dziś ludzie polegają na doświadczeniu, intuicji i wewnętrznych statystykach, aby ocenić przyszłe wydarzenia, tak w nadchodzących latach przyszłość projektów budowlanych będzie w coraz większym stopniu określana przez połączenie zgromadzonej wiedzy i matematycznych modeli uczenia maszynowego.



Rys. 9.2-12 Jakościowe i ustrukturyzowane dane historyczne firmy są materiałem, na którym budowane są modele uczenia maszynowego i prognozy.

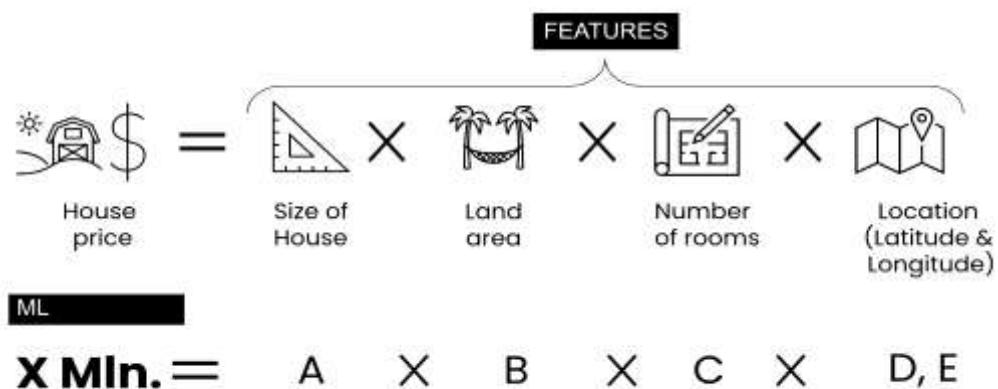
Rozważmy prosty przykład: przewidywanie ceny domu na podstawie jego powierzchni, wielkości działki, liczby pokoi i położenia geograficznego. Jednym z podejść jest zbudowanie klasycznego modelu, który analizuje te parametry i oblicza oczekiwana cenę (Rys. 9.2-13). Podejście to wymaga precyzyjnego i znanego z góry wzoru, co jest praktycznie niemożliwe w praktyce.



Rys. 9.2-13 Klasyczny algorytm może być użyty do oszacowania wartości domu z ustalonym wzorem do znalezienia.

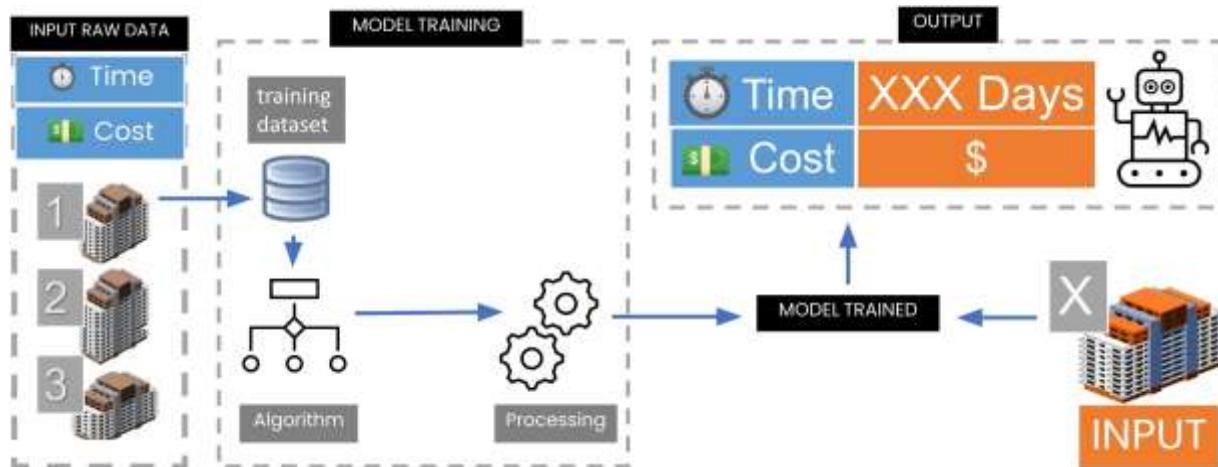
Uczenie maszynowe eliminuje ręczne wyszukiwanie formuł i zastępuje je wyszkolonymi algorytmami, które niezależnie identyfikują zależności, które są wielokrotnie dokładniejsze niż jakiekolwiek wcześniej ustalone równania. Alternatywnie, stwórzmy algorytm uczenia maszynowego, który wygeneruje model na podstawie wcześniejszego zrozumienia problemu i danych historycznych, które mogą być niekompletne (Rys. 9.2-14).

Na przykładzie wyceny, uczenie maszynowe umożliwia tworzenie różnego rodzaju modeli matematycznych, które nie wymagają znajomości dokładnego mechanizmu powstawania kosztów. Model "uczy się" na podstawie danych dotyczących poprzednich projektów, dostosowując się do rzeczywistych wzorców między parametrami budynku, ich kosztami i terminami.



Rys. 9.2-14 W przeciwieństwie do klasycznej estymacji opartej na formułach, algorytm uczenia maszynowego jest szkolony na danych historycznych.

W kontekście nadzorowanego uczenia maszynowego, każdy projekt w szkoleniowym zbiorze danych zawiera zarówno atrybuty wejściowe (np. koszt i czas podobnych budynków), jak i oczekiwane wartości wyjściowe (np. koszt lub czas). Podobny zbiór danych jest wykorzystywany do tworzenia i dostosowywania modelu uczenia maszynowego (Rys. 9.2-15). Im większy zbiór danych i im wyższa jakość danych w nim zawartych, tym dokładniejszy będzie model i tym dokładniejsze będą wyniki przewidywania.



Rys. 9.2-15 Model ML wyszkolony na danych dotyczących kosztów i harmonogramu z poprzednich projektów określi koszt i harmonogram nowego projektu z pewnym prawdopodobieństwem.

Po utworzeniu modelu i przeszkoleniu go w zakresie szacowania budowy nowego projektu, wystarczy dostarczyć modelowi nowe atrybuty dla nowego projektu, a model dostarczy szacunkowe wyniki w oparciu o wcześniej wyuczone wzorce z pewnym prawdopodobieństwem.

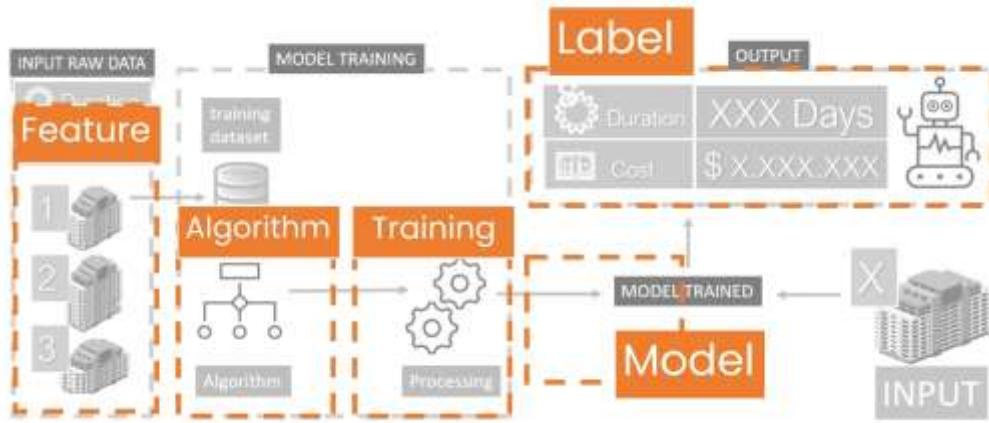
Kluczowe koncepcje uczenia maszynowego

Uczenie maszynowe to nie magia, to tylko matematyka, dane i znajdowanie wzorców. Nie ma prawdziwej inteligencji, ale jest programem przeszkolonym w zakresie danych w celu rozpoznawania wzorców i

podejmowania decyzji bez ciągłego zaangażowania człowieka.

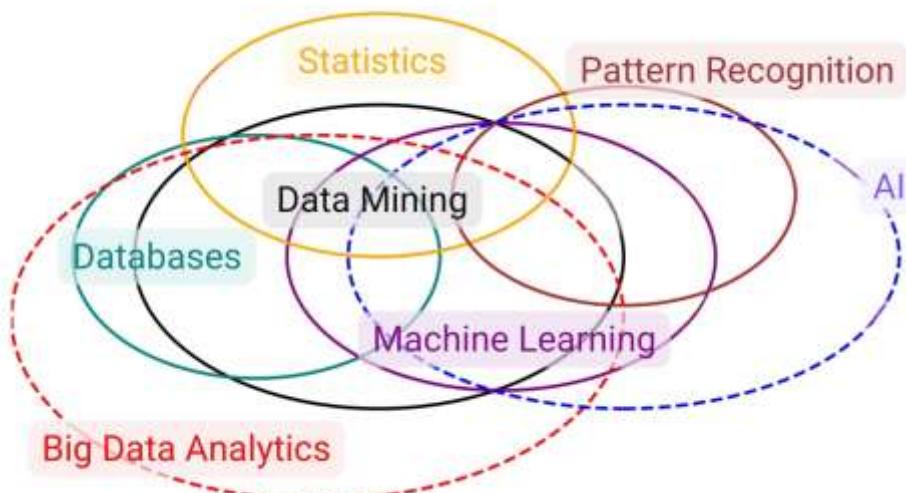
Uczenie maszynowe wykorzystuje szereg kluczowych pojęć do opisania swojej struktury (Rys. 9.2-16):

- **Etykiety** to docelowe zmienne lub atrybuty (parametr "Survivor" w zbiorze danych Titanica), które model powinien przewidywać. Przykład: koszt budowy (np. w dolarach), czas trwania prac budowlanych (np. w miesiącach).
- **Cechy** to niezależne zmienne lub atrybuty, które służą jako dane wejściowe do modelu. W modelu prognozowania są one wykorzystywane do przewidywania etykiet. Przykłady: wielkość działki (w metrach kwadratowych), liczba pięter budynku, całkowita powierzchnia budynku (w metrach kwadratowych), położenie geograficzne (szerokość i długość geograficzna), rodzaj materiałów użytych w budownictwie. Liczba cech określa również wymiarowość danych.
- **Model** to zestaw różnych hipotez, z których jedna przybliża funkcję docelową, która ma być przewidywana lub przybliżana. Przykład: model uczenia maszynowego, który wykorzystuje techniki analizy regresji do przewidywania kosztów i harmonogramu budowy.
- **Algorytm uczenia** **Algorytm uczenia** to proces znajdowania najlepszej hipotezy w modelu, która dokładnie pasuje do funkcji docelowej przy użyciu zestawu danych szkoleniowych. Przykład: Algorytm regresji liniowej, KNN lub losowego lasu, który analizuje dane dotyczące kosztów i czasu budowy w celu zidentyfikowania relacji i wzorców.
- **Szkolenie** - podczas procesu szkolenia algorytm analizuje dane szkoleniowe, znajdując wzorce, które odpowiadają relacji między atrybutami wejściowymi a etykietami docelowymi. Wynikiem tego procesu jest wytrenowany model uczenia maszynowego, gotowy do przewidywania. Przykład: proces, w którym algorytm analizuje historyczne dane budowlane (koszt, czas, charakterystyka obiektu) w celu stworzenia modelu predykcyjnego.



Rys. 9.2-16 ML wykorzystuje etykiety i atrybuty do tworzenia modeli, które są trenowane na danych przy użyciu algorytmów do przewidywania wyników.

Uczenie maszynowe nie istnieje w izolacji, ale jest częścią szerszego ekosystemu dyscyplin analitycznych, w tym statystyki, baz danych, eksploracji danych, rozpoznawania wzorców, analizy dużych zbiorów danych i sztucznej inteligencji. Rys. 9.2-17 pokazuje, w jaki sposób te dziedziny nakładają się i uzupełniają, zapewniając kompleksowe ramy dla nowoczesnych systemów podejmowania decyzji i automatyzacji.

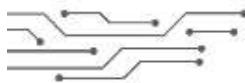


Rys. 9.2-17 Związek między różnymi obszarami analizy danych: statystyką, uczeniem maszynowym, sztuczną inteligencją, big data, rozpoznawaniem wzorców i eksploracją danych.

Główne celem uczenia maszynowego jest wyposażenie komputerów w zdolność do automatycznego uczenia się wiedzy bez interwencji lub pomocy człowieka i odpowiedniego dostosowywania swoich działań [159].

Tak więc w przyszłości rolą człowieka będzie jedynie zapewnienie maszynie zdolności poznawczych - będzie on ustalał warunki, wagi i parametry, a model uczenia maszynowego zajmie się resztą.

W kolejnym rozdziale rozważymy konkretne przykłady zastosowania algorytmu. Rzeczywiste tabele i uproszczone modele zostaną wykorzystane do pokazania, jak prognoza jest budowana krok po kroku.



ROZDZIAŁ 8.3.

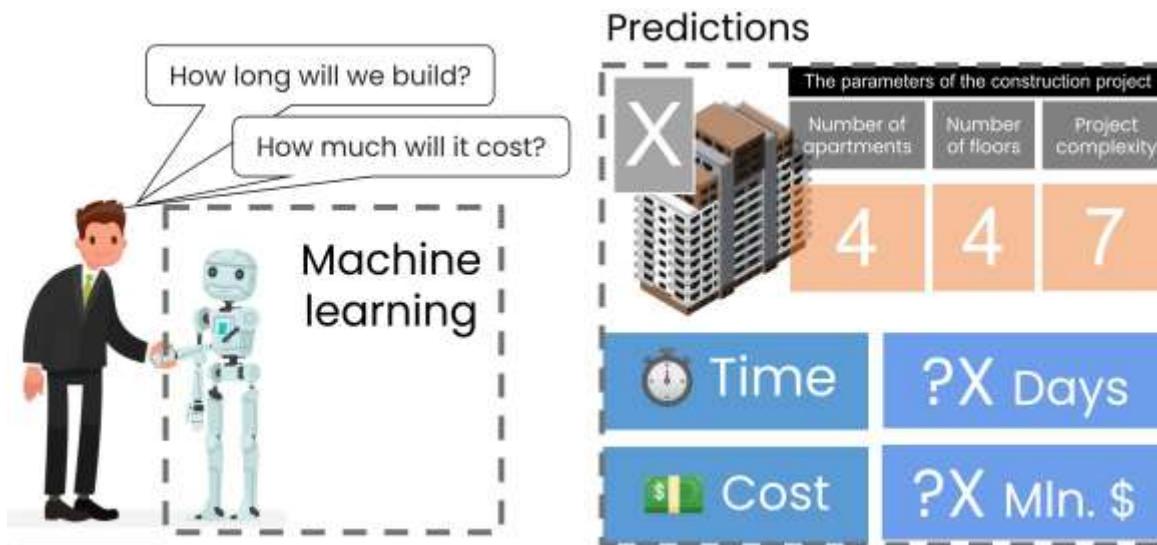
PRZEWIDYWANIE KOSZTÓW I TERMINÓW PRZY UŻYCIU UCZENIA MASZYNOWEGO

Przykład wykorzystania uczenia maszynowego do znalezienia kosztów i harmonogramów projektu

Szacowanie czasu i kosztów budowy jest jednym z kluczowych procesów w działalności firmy budowlanej. Tradycyjnie takie szacunki były dokonywane przez ekspertów w oparciu o doświadczenie, podręczniki i regulacyjne bazy danych. Jednak wraz z transformacją cyfrową i rosnącą dostępnością danych, możliwe jest obecnie wykorzystanie modeli uczenia maszynowego (ML) w celu poprawy dokładności i automatyzacji takich szacunków.

Wprowadzenie uczenia maszynowego do procesu obliczania kosztów i harmonogramu budowy nie tylko pozwala na bardziej efektywne planowanie, ale także staje się punktem wyjścia do integracji inteligentnych modeli z innymi procesami biznesowymi - od zarządzania ryzykiem po optymalizację logistyki i zaopatrzenia.

Ważne jest, aby móc szybko określić, ile czasu zajmie realizacja projektu i jaki będzie jego całkowity koszt. Te pytania dotyczące czasu i kosztów projektu tradycyjnie znajdują się w czołówce umysłów zarówno klientów, jak i firm budowlanych od momentu narodzin branży budowlanej.



Rys. 9.3-1 W projektach budowlanych szybkość i jakość szacowania czasu i kosztów budowy są kluczowymi czynnikami sukcesu.

W poniższym przykładzie kluczowe dane z poprzednich projektów zostaną wyodrębnione i wykorzystane do opracowania modelu uczenia maszynowego, który pozwoli nam wykorzystać model do oszacowania kosztów i harmonogramu nowych projektów budowlanych z nowymi parametrami (Rys. 9.3-1).

Rozważmy trzy projekty z trzema kluczowymi atrybutami: liczbą mieszkań (gdzie 100 mieszkań jest równoważne liczbie 10 dla ułatwienia wizualizacji), liczbą pięter i warunkową miarą złożoności konstrukcji w skali od 1 do 10, gdzie 10 to najwyższy wynik złożoności. W uczeniu maszynowym proces konwersji i upraszczania wartości takich jak 100 do 10 lub 50 do 5 nazywany jest "normalizacją".

Normalizacja w uczeniu maszynowym to proces sprowadzania różnych danych liczbowych do wspólnej skali w celu ułatwienia przetwarzania i analizy. Proces ten jest szczególnie ważny, gdy dane mają różne skale i jednostki miary.

Załóżmy, że pierwszy projekt (Rys. 9.3-2) miał 50 mieszkań (po normalizacji 5), 7 kondygnacji i wynik złożoności 2, co oznaczało stosunkowo prostą konstrukcję. Drugi projekt miał już 80 mieszkań, 9 kondygnacji i stosunkowo złożony projekt. W tych warunkach budowa pierwszego i drugiego budynku mieszkalnego zajęła odpowiednio 270 i 330 dni, a całkowity koszt projektu wyniósł odpowiednio 4,5 mln USD i 5,8 mln USD.

The parameters of the construction project			
Construction project	Number of apartment	Number of floors	Project complexity
1	5	7	2
2	8	9	6
3	3	5	3
X	4	4	7

The key parameters of the project	
Time	Cost
270	\$ 4.502.000
330	\$ 5.750.000
230	\$ 3.262.000
?X	\$?X. XXX.XXX

Rys. 9.3-2 Przykład zestawu przeszłych projektów, które zostaną wykorzystane do oszacowania czasu i kosztów przyszłego projektu X.

Podczas budowania modelu uczenia maszynowego dla takich danych, głównym zadaniem jest zidentyfikowanie krytycznych atrybutów (lub etykiet) do przewidywania, w tym przypadku czasu i kosztów budowy. Przy niewielkim zbiorze danych wykorzystamy informacje o poprzednich projektach budowlanych do planowania nowych: korzystając z algorytmów uczenia maszynowego, musimy przewidzieć koszt i czas trwania budowy nowego projektu X w oparciu o dane atrybuty nowego projektu, takie jak 40 mieszkań, 4 piętra i względnie wysoka złożoność projektu wynosząca 7 (rys. 9.3-2). W warunkach rzeczywistych liczba parametrów wejściowych może być znacznie większa i wynosić od kilkudziesięciu do kilkuset czynników. Mogą one obejmować: rodzaj materiałów budowlanych, strefę klimatyczną, poziom kwalifikacji wykonawców, dostępność mediów, rodzaj fundamentów, sezon rozpoczęcia prac, komentarze brygadzistów itp.

Aby stworzyć predykcyjny model uczenia maszynowego, musimy wybrać algorytm do jego utworzenia. Algorytm w uczeniu maszynowym jest jak przepis matematyczny, który uczy komputer, jak przewidywać (mieszać w odpowiedniej kolejności parametry) lub podejmować decyzje na podstawie danych.

Do analizy danych dotyczących wcześniejszych projektów budowlanych i przewidywania czasu i kosztów przyszłych projektów (Rys. 9.3-2) można wykorzystać jeden z popularnych algorytmów uczenia maszynowego:

- **Regresja liniowa (regresja liniowa):** algorytm ten próbuje znaleźć bezpośredni związek między atrybutami, na przykład między liczbą pięter a kosztem budowy. Celem algorytmu jest znalezienie równania liniowego, które najlepiej opisuje ten związek, co pozwala na prognozowanie.
- **Algorytm k-najbliższych sąsiadów (k-NN):** algorytm ten porównuje nowy projekt z wcześniejszymi projektami o podobnym rozmiarze lub złożoności. K-NN klasyfikuje dane w oparciu o to, które z k (liczby) przykładów szkoleniowych są do nich najbliższe. W kontekście regresji wynikiem jest średnia lub mediana k najbliższych sąsiadów.
- **Drzewa decyzyjne:** to model modelowania predykcyjnego, który dzieli dane na podzbiory w oparciu o różne warunki przy użyciu struktury drzewa. Każdy węzeł drzewa reprezentuje warunek lub pytanie prowadzące do dalszego podziału danych, a każdy liść reprezentuje ostateczną prognozę lub wynik. Algorytm dzieli dane na mniejsze grupy w oparciu o różne cechy, takie jak najpierw liczba historii, następnie złożoność i tak dalej, aby dokonać prognozy.

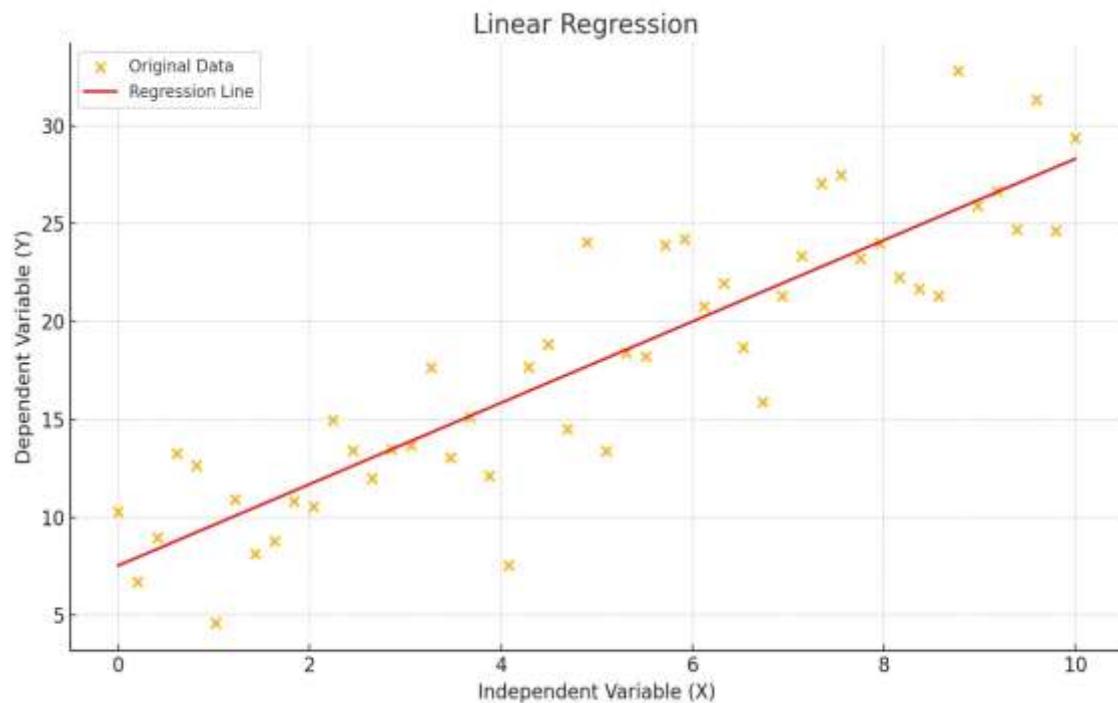
Przyjrzyjmy się algorytmom uczenia maszynowego do szacowania kosztów nowego projektu na przykładzie dwóch popularnych algorytmów: regresji liniowej i algorytmu K-najbliższych sąsiadów.

Przewidywanie kosztów i czasu projektu przy użyciu regresji liniowej

Regresja liniowa to podstawowy algorytm analizy danych, który przewiduje wartość zmiennej na podstawie liniowej zależności z jedną lub większą liczbą innych zmiennych. Model ten zakłada, że istnieje bezpośrednia liniowa zależność między zmienną zależną a jedną lub większą liczbą zmiennych niezależnych, a celem algorytmu jest znalezienie tej zależności.

Prostota i przejrzystość regresji liniowej sprawiły, że stała się ona popularnym narzędziem w wielu dziedzinach. W przypadku pojedynczej zmiennej regresja liniowa polega na znalezieniu najlepiej dopasowanej linii przechodzącej przez punkty danych.

Regresja liniowa znajduje najlepszą linię prostą (czerwona linia), która przybliża zależność między zmienną wejściową X i zmienną wyjściową Y. Linia ta umożliwia przewidywanie wartości Y dla nowych wartości X w oparciu o zidentyfikowaną zależność liniową (Rys. 9.3-3).



Rys. 9.3-3 Zasada regresji liniowej polega na znalezieniu najlepszej linii prostej przechodzącej przez wartości treningowe.

Linia ta jest reprezentowana przez równanie, w którym wprowadzenie wartości zmiennej niezależnej (X) daje przewidywaną wartość zmiennej zależnej (Y). Proces ten pozwala na skuteczne przewidywanie Y na podstawie znanych wartości X przy użyciu liniowej zależności między nimi. Przykład znalezienia takiej statystycznie uśrednionej linii można zobaczyć w ocenie danych dotyczących pozwoleń na budowę w San Francisco (Rys. 9.1-7), gdzie inflacja przy użyciu regresji liniowej została obliczona dla różnych typów obiektów.

Załadujmy tabelę danych projektu (Rys. 9.3-2 z poprzedniego rozdziału) bezpośrednio do programu LLM i poprośmy go o zbudowanie dla nas prostego modelu uczenia maszynowego.

- ✍ Wyślij żądanie tekstowe do czatu LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN:

Musimy pokazać budowę prostego modelu uczenia maszynowego do przewidywania kosztów i czasu realizacji nowego projektu X (rys. 9.3-2 jako załączony obraz) ↴

2 Odpowiedź LLM :

```

1  from sklearn.linear_model import LinearRegression
2  import numpy as np
3
4  # Data from the image
5  # Features: Number of apartments, number of floors, project complexity
6  X = np.array([
7      [5, 7, 2],
8      [8, 9, 6],
9      [3, 5, 3]
10 ])
11
12 # Targets: Time (days), Cost ($)
13 y_time = np.array([270, 330, 230])
14 y_cost = np.array([4502000, 5750000, 3262000])
15
16 # Creating linear regression models for time and cost
17 model_time = LinearRegression()
18 model_cost = LinearRegression()
19
20 # Training the models
21 model_time.fit(X, y_time)
22 model_cost.fit(X, y_cost)
23
24 # The new project's features
25 X_new = np.array([[4, 4, 7]])
26
27 # Predicting the time and cost for the new project
28 predicted_time = model_time.predict(X_new)
29 predicted_cost = model_cost.predict(X_new)
30
31 predicted_time, predicted_cost

```

Result
(array([238.44444444]), array([3042337.77777778]))

Rys. 9.3-4 LLM wybrał regresję liniową do stworzenia modelu uczenia maszynowego w celu przewidywania kosztów i czasu projektu.

Program LLM automatycznie rozpoznał tabelę z załączonego obrazu i przekonwertował dane z formatu wizualnego na tablicę (rys. 9.3-4 - wiersz 6). Tablica ta została wykorzystana jako podstawa do utworzenia cech i etykiet, na podstawie których utworzono model uczenia maszynowego (rys. 9.3-4 - 17-22 wiersz), który wykorzystywał regresję liniową.

Korzystając z podstawowego modelu regresji liniowej, który został wytrenowany na "niezwykle małym" zbiorze danych, dokonano prognoz dla nowego hipotetycznego projektu budowlanego oznaczonego jako Projekt X. W naszym problemie projekt ten charakteryzuje się 40 mieszkańami, 4 piętrami i poziomem

złożoności 7 (Rys. 9.3-2).

Zgodnie z przewidywaniami modelu regresji liniowej opartego na ograniczonym i niewielkim zestawie danych dla nowego Projektu X (Rys. 9.3-4 - linia 24-29):

- **Czas trwania budowy wyniesie** około 238 dni (238.4444444).
- **Całkowite wydatki wyniosą** w przybliżeniu \$ 3,042,338 (3042337.777).

Aby dokładniej zbadać hipotezę dotyczącą kosztów projektu, warto poeksperymentować z różnymi algorytmami i metodami uczenia maszynowego. Dlatego przewidujemy te same wartości kosztów i czasu dla nowego projektu X w oparciu o niewielki zestaw danych historycznych przy użyciu algorytmu K-Najbliższych Sąsiadów (k-NN).

Prognozy kosztów i czasu projektu przy użyciu algorytmu K-najbliższych sąsiadów (k-NN)

Używamy algorytmu k-Nearest Neighbours (k-NN) jako dodatkowego predyktora do oszacowania kosztów i czasu trwania nowego projektu. Algorytm K-Nearest Neighbours (k-NN) jest nadzorowaną metodą uczenia maszynowego (nadzorowanego uczenia maszynowego) zarówno dla klasyfikacji, jak i regresji. Algorytm k-NN omówiliśmy również wcześniej w kontekście przeszukiwania wektorowej bazy danych (Rys. 8.2-2), gdzie jest on używany do znajdowania najbliższych wektorów (np. tekstów, obrazów lub opisów technicznych). W tym podejściu każdy projekt jest reprezentowany jako punkt w wielowymiarowej przestrzeni, gdzie każdy wymiar odpowiada określonemu atrybutowi projektu.

W naszym przypadku, biorąc pod uwagę trzy atrybuty każdego projektu, przedstawimy je jako punkty w trójwymiarowej przestrzeni (Rys. 9.3-5). Tak więc nasz nadchodzący projekt X będzie zlokalizowany w tej przestrzeni o współrzędnych ($x=4$, $y=4$, $z=7$). Należy zauważyć, że w warunkach rzeczywistych liczba punktów i wymiarowość przestrzeni mogą być o rzędy wielkości większe.

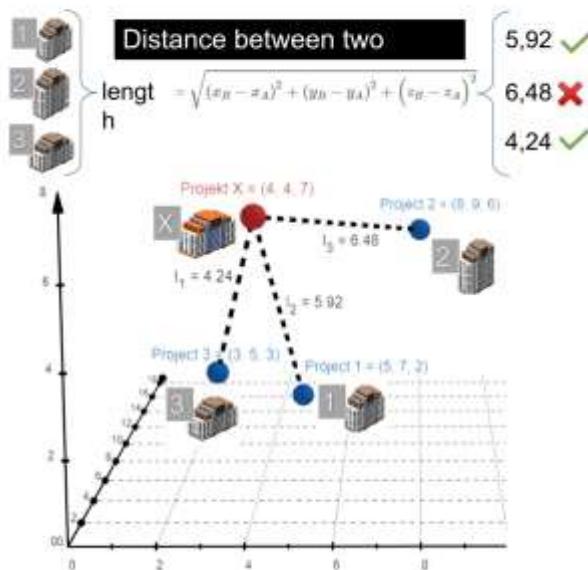
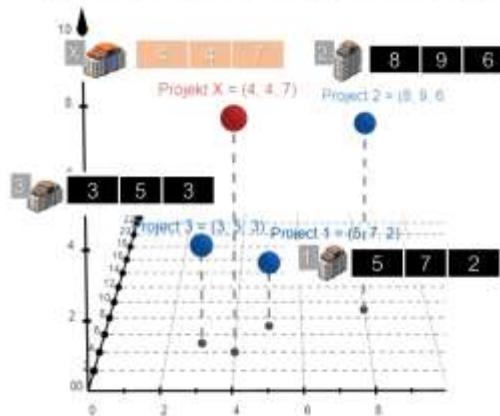
Algorytm K-NN (k-najbliższych sąsiadów) działa poprzez pomiar odległości między pożądanym projektem X a projektami w szkoleniowej bazie danych. Porównując te odległości, algorytm określa projekty, które znajdują się najbliżej punktu nowego projektu X.

Na przykład, jeśli drugi projekt ($x=8$, $y=9$, $z=6$) z naszego oryginalnego zbioru danych znajduje się znacznie dalej od X (Rys. 9.3-5) niż pozostałe projekty, można go wykluczyć z dalszej analizy. W rezultacie do obliczeń można wykorzystać tylko dwa ($k=2$) najbliższe projekty, na podstawie których zostanie określona wartość średnia.

Taka metoda, poprzez wyszukiwanie sąsiedztwa, pozwala ocenić podobieństwa między projektami, co z kolei pomaga wyciągnąć wnioski na temat prawdopodobnych kosztów i harmonogramu nowego projektu w oparciu o podobne projekty, które zostały wdrożone wcześniej.

k-nearest neighbors algorithm

The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other.



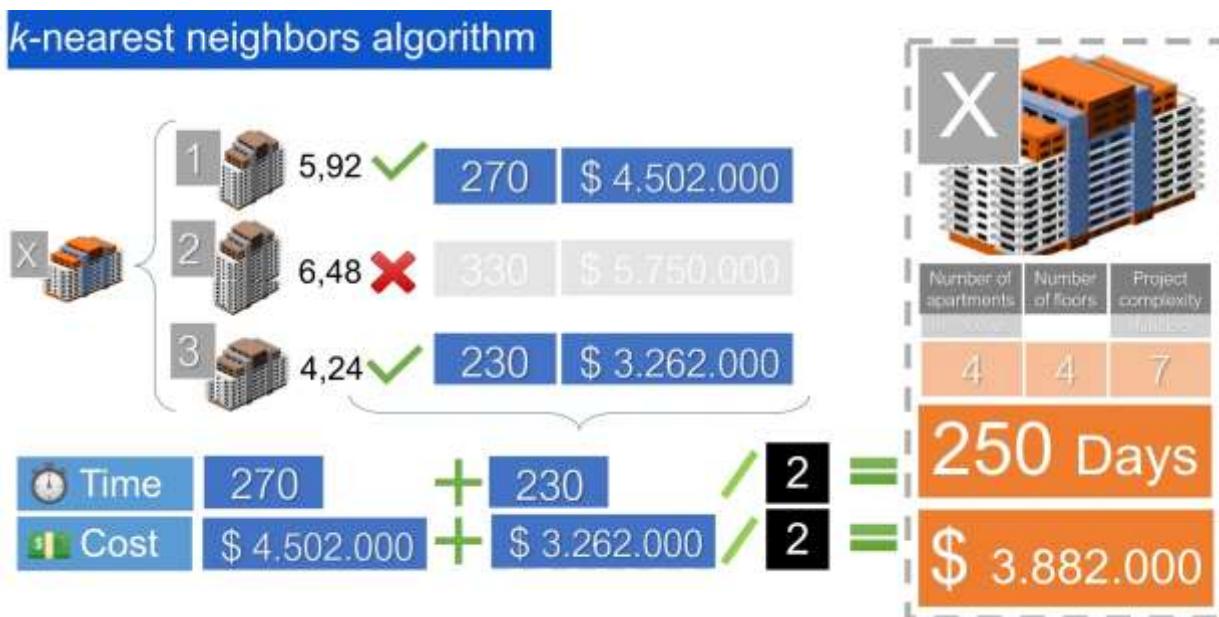
Rys. 9.3-5 W algorytmie K-NN projekty są reprezentowane jako punkty w wielowymiarowej przestrzeni, a najbliższe projekty są wybierane na podstawie odległości do oceny podobieństwa i przewidywania.

Praca k-NN obejmuje kilka kluczowych kroków:

- **Przygotowanie danych:** najpierw ładowane są zestawy danych treningowych i testowych. Dane treningowe służą do "wytrenowania" algorytmu, a dane testowe do sprawdzenia jego skuteczności.
- **Wybór parametru K:** wybierana jest liczba K, która wskazuje, ilu najbliższych sąsiadów (punktów danych) powinno zostać uwzględnionych w algorytmie. Wartość "K" jest bardzo ważna, ponieważ wpływa na wynik.
- Proces klasyfikacji i regresja dla danych testowych:
 - **Obliczanie odległości:** dla każdego elementu z danych testowych obliczana jest odległość do każdego elementu z danych treningowych (Rys. 9.3-5). W tym celu można zastosować różne metody pomiaru odległości, takie jak odległość euklidesowa (najczęściej stosowana metoda), odległość Manhattan lub odległość Hamminga.
 - **Sortowanie i wybór K najbliższych sąsiadów:** po obliczeniu odległości są one sortowane i wybieranych jest K punktów najbliższych punktowi testowemu.
 - **Określanie klasy lub wartości punktu** testowego: jeśli jest to zadanie klasyfikacji, klasa punktu testowego jest określana na podstawie najczęściej występującej klasy wśród K wybranych sąsiadów. Jeśli jest to zadanie regresji, obliczana jest średnia (lub inna miara tendencji centralnej) wartości K sąsiadów.
- **Zakończenie** procesu: po sklasyfikowaniu wszystkich danych testowych lub dokonaniu dla nich prognoz, proces jest zakończony.

Algorytm k-najbliższych sąsiadów (k-NN) jest skuteczny w wielu praktycznych zastosowaniach i jest jednym z głównych narzędzi w arsenale specjalistów od uczenia maszynowego. Algorytm ten jest popularny ze względu na swoją prostotę i wydajność, szczególnie w zadaniach, w których relacje między danymi są łatwe do zinterpretowania.

W naszym przykładzie, po zastosowaniu algorytmu K-najbliższych sąsiadów, zidentyfikowano dwa projekty (z naszej małej próbki) o najmniejszej odległości od projektu X (Rys. 9.3-5). Na podstawie tych projektów algorytm określa średnią ich cenę i czas trwania budowy. Po analizie (rys. 9.3-6) algorytm, uśredniając najbliższych sąsiadów, doszedł do wniosku, że projekt X będzie kosztował około \$ 3 800 000 USD, a jego ukończenie zajmie około 250 dni.



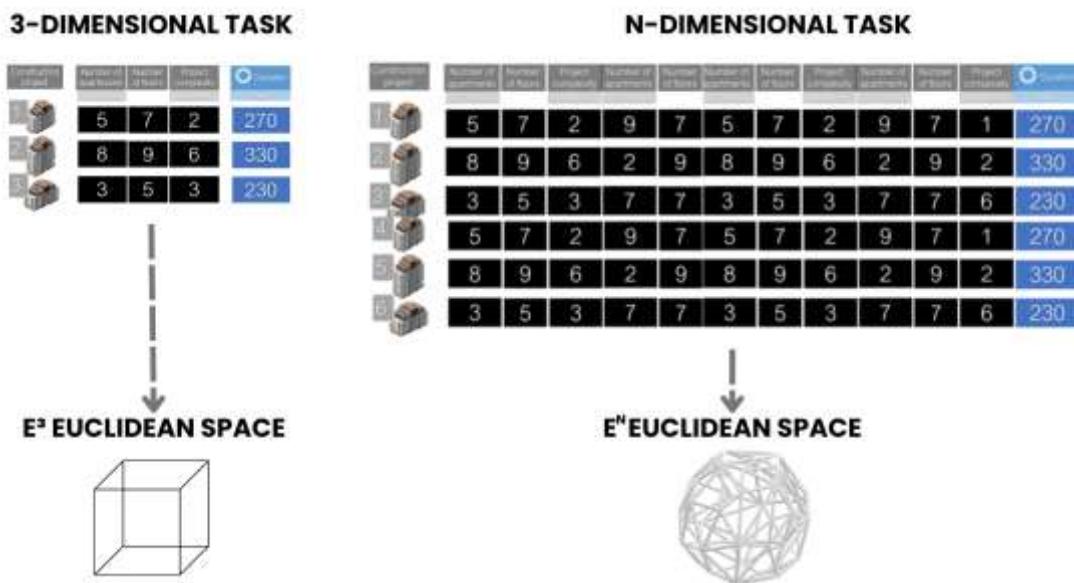
Rys. 9.3-6 Algorytm K-najbliższych sąsiadów określa koszt i harmonogram projektu X poprzez analizę dwóch najbliższych projektów w próbie.

Algorytm k-Nearest Neighbors (k-NN) jest szczególnie popularny w zadaniach klasyfikacji i regresji, takich jak systemy rekomendacji, gdzie jest wykorzystywany do sugerowania produktów lub treści w oparciu o preferencje podobne do zainteresowań konkretnego użytkownika. Ponadto k-NN jest szeroko stosowany w diagnostyce medycznej do klasyfikowania rodzajów chorób na podstawie objawów pacjenta, w rozpoznawaniu wzorców oraz w sektorze finansowym do oceny zdolności kredytowej klientów.

Nawet przy ograniczonej ilości danych, modele uczenia maszynowego mogą zapewnić użyteczne prognozy i znacząco poprawić analityczny komponent zarządzania projektami budowlanymi. W miarę rozszerzania i oczyszczania danych historycznych możliwe jest przejście do bardziej wyrafinowanych modeli - na przykład uwzględniających rodzaj budowy, lokalizację, sezon rozpoczęcia budowy i inne czynniki.

Nasze uproszczone zadanie wykorzystywało trzy atrybuty do wizualizacji w przestrzeni 3D, ale rzeczywiste

projekty zawierają średnio setki lub tysiące atrybutów (patrz zbiór danych z rozdziału "Przykład big data opartego na CAD (BIM)"), co znacznie zwiększa wymiarowość przestrzeni i złożoność reprezentowania projektów jako wektorów (Rys. 9.3-7).



Rys. 9.3-7 W uproszczonym przykładzie do wizualizacji 3D użyto trzech atrybutów, podczas gdy rzeczywiste projekty mają ich więcej.

Zastosowanie różnych algorytmów do tego samego zbioru danych dla projektu X, który ma 40 mieszkań, 4 piętra i poziom złożoności 7, dało różne przewidywane wartości. Algorytm regresji liniowej przewidywał czas ukończenia wynoszący 238 dni i koszt \$ 3 042 338 (Rys. 9.3-4), podczas gdy algorytm k-NN przewidywał 250 dni i \$ 3 882 000 (Rys. 9.3-6).

Dokładność przewidywań uzyskanych przy użyciu modeli uczenia maszynowego, zależy bezpośrednio od ilości i jakości danych wejściowych. Im więcej projektów jest zaangażowanych w szkolenie oraz im pełniej i dokładniej reprezentowane są ich cechy (atrybuty) i wyniki (etykiety), tym większe prawdopodobieństwo uzyskania wiarygodnych prognoz przy minimalnych wartościach błędów.

Techniki wstępnego przetwarzania danych odgrywają ważną rolę w tym procesie:

- Normalizacja w celu sprowadzenia cech do wspólnej skali;
- Wykrywanie i eliminacja wartości odstających, eliminacja zniekształceń modelu;
- Kodowanie atrybutów kategorycznych w celu umożliwienia manipulacji danymi tekstowymi;
- Uzupełnianie brakujących wartości, zwiększające odporność modelu.

Ponadto metody walidacji krzyżowej są wykorzystywane do oceny ogólności modelu i jego odporności na nowe zbiory danych w celu wykrycia nadmiernego dopasowania i poprawy wiarygodności prognoz.

Chaos jest porządkiem, który należy rozszyfrować [160].

- José Saramago, "The Double"

Nawet jeśli wydaje ci się, że chaosu twoich zadań nie da się opisać formalnie, powinieneś wiedzieć, że każde wydarzenie na świecie, a zwłaszcza procesy budowlane, podlegają prawom matematycznym, które mogą wymagać wsparcia obliczania wartości nie za pomocą ścisłych formuł, ale za pomocą statystyk i danych historycznych.

Zarówno tradycyjne szacunki działań kosztorysowych, jak i modele uczenia maszynowego nieuchronnie napotykają niepewność i potencjalne źródła błędów. Jeśli jednak dostępne są dane o wystarczającej jakości, modele uczenia maszynowego mogą wykazywać porównywalną, a czasem nawet wyższą dokładność przewidywania niż oceny ekspertów.

Uczenie maszynowe prawdopodobnie stanie się niezawodnym narzędziem uzupełniającym analizy, które może: udoskonalać obliczenia, sugerować alternatywne scenariusze i identyfikować ukryte zależności między parametrami projektu. Takie modele nie będą rościć sobie prawa do uniwersalności, ale wkrótce zajmą ważne miejsce w obliczeniach i procesach decyzyjnych. Technologie uczenia maszynowego nie wykluczą udziału inżynierów, estymatorów i analityków, a wręcz przeciwnie, rozszerzą ich możliwości, oferując dodatkowy punkt widzenia oparty na danych historycznych.

Odpowiednio zintegrowane z procesami biznesowymi firm budowlanych, uczenie maszynowe ma potencjał, aby stać się ważnym elementem systemu wspomagania decyzji zarządczych - nie jako zastępstwo dla ludzi, ale jako rozszerzenie ich profesjonalnej intuicji i logiki inżynierskiej.

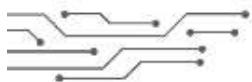
Kolejne kroki: od przechowywania do analizy i prognozowania

Nowoczesne podejście do pracy z danymi zaczyna zmieniać sposób podejmowania decyzji w branży budowlanej. Przejście od intuicyjnych ocen do obiektywnej analizy danych nie tylko poprawia dokładność, ale także otwiera nowe możliwości optymalizacji procesów. Podsumowując tę część, warto podkreślić główne praktyczne kroki, które pomogą zastosować omawiane metody w codziennych zadaniach:

- Ustanowienie zrównoważonej infrastruktury magazynowania
 - Próba połączenia różnych dokumentów i danych projektowych w jeden model tabelaryczny, agregujący kluczowe informacje w jedną ramkę danych do dalszej analizy.
 - Używaj wydajnych formatów przechowywania danych - na przykład formatów kolumnowych, takich jak Apache Parquet zamiast CSV lub XLSX - szczególnie w przypadku tych zestawów, które mogą być potencjalnie wykorzystane do trenowania modeli uczenia maszynowego w przyszłości.
 - Ustanowienie systemu wersjonowania danych w celu śledzenia zmian w całym projekcie.
- Wdrażanie narzędzi do analizy i automatyzacji

- Rozpoczęcie analizy historycznych danych projektu - według dokumentacji, modeli, szacunków - w celu zidentyfikowania wzorców, trendów i anomalii.
 - Opanowanie procesów ETL (Extract, Transform, Load) w celu automatycznego ładowania i przygotowywania danych.
 - Dowiedz się, jak wizualizować kluczowe metryki przy użyciu różnych bezpłatnych bibliotek wizualizacji Pythona.
 - Rozpoczęcie stosowania metod statystycznych i losowego pobierania próbek w celu uzyskania reprezentatywnych i powtarzalnych wyników analitycznych.
- Rosnąca dojrzałość w pracy z danymi
- Poznaj kilka podstawowych algorytmów uczenia maszynowego na prostych i nieskomplikowanych przykładach, takich jak zbiór danych Titanic.
 - Analizować bieżące procesy i identyfikować obszary, w których można odejść od sztywnej logiki przyczynowo-skutkowej na rzecz statystycznych metod prognozowania i szacowania.
 - Zacznią traktować dane jako zasób strategiczny, a nie produkt uboczny: buduj procesy decyzyjne wokół modeli danych, a nie wokół konkretnych rozwiązań programowych.

Firmy budowlane, które zdały sobie sprawę z wartości danych, wkraczają w nową fazę rozwoju, w której o przewadze konkurencyjnej decyduje nie ilość zasobów, ale szybkość podejmowania decyzji w oparciu o analitykę.



MAKSYMALNA WYGODA DZIĘKI WERSJI DRUKOWANEJ

Posiadasz darmową wersję cyfrową **Data-Driven Construction**. W celu wygodniejszej pracy i szybkiego dostępu do materiałów zalecamy zwrócenie uwagi na **wydanie drukowane**:



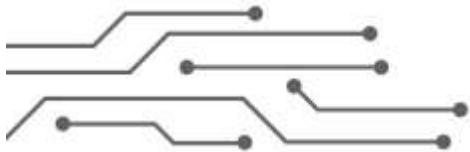
■ **Zawsze pod ręką:** książka w formacie drukowanym stanie się niezawodnym narzędziem pracy, pozwalającym szybko znaleźć i wykorzystać odpowiednie wizualizacje i diagramy w każdej sytuacji roboczej.

■ **Wysokiej jakości ilustracje:** wszystkie obrazy i grafiki w drukowanym wydaniu są prezentowane w maksymalnej jakości.

■ **Szybki dostęp do informacji:** łatwa nawigacja, możliwość tworzenia notatek, zakładek i pracy z książką w dowolnym miejscu.

Kupując pełną drukowaną wersję książki, otrzymujesz wygodne narzędzie do wygodnej i efektywnej pracy z informacją: możliwość szybkiego korzystania z materiałów wizualnych w codziennych zadaniach, szybkiego wyszukiwania potrzebnych schematów i robienia notatek. Ponadto zakup wspiera rozpowszechnianie otwartej wiedzy.

Zamów drukowaną wersję książki na stronie: datadrivenconstruction.io/books



IX CZĘŚĆ **BRANŻA BUDOWLANA W ERZE CYFROWEJ. MOŻLIWOŚCI I WYZWANIA**

Ostatnia, dziesiąta część to kompleksowe spojrzenie na przyszłość branży budowlanej w erze cyfrowej transformacji. Analizuje przejście od analizy przyczynowej do pracy z korelacjami dużych zbiorów danych. Zarysowano podobieństwa między ewolucją sztuk pięknych a rozwojem pracy z danymi w budownictwie, pokazując, jak branża przechodzi od szczegółowej kontroli do holistycznego zrozumienia procesów. Badana jest koncepcja "uberyzacji" branży budowlanej, w której przejrzystość danych i zautomatyzowane obliczenia mogą radykalnie zmienić tradycyjne modele biznesowe, eliminując potrzebę pośredników i ograniczając możliwości spekulacji. Nierozwiążane kwestie, takie jak uniwersalna klasyfikacja elementów, zostały szczegółowo omówione, dając firmom budowlanym czas na dostosowanie się do nowego środowiska. Część kończy się konkretnymi zaleceniami dotyczącymi kształtowania strategii transformacji cyfrowej, która obejmuje analizę słabych punktów i rozszerzanie usług w celu utrzymania konkurencyjności w zmieniającej się branży.

ROZDZIAŁ 9.1.

STRATEGIE PRZETRWANIA: BUDOWANIE PRZEWAGI KONKURENCYJNEJ

Korelacje zamiast obliczeń: przyszłość analityki budowlanej

Ze względu na szybką cyfryzację informacji (Rys. 1.1-5), nowoczesne budownictwo przechodzi fundamentalną transformację, w której dane stają się nie tylko narzędziem, ale strategicznym zasobem, który może zasadniczo zmienić tradycyjne podejście do zarządzania projektami i biznesem.

Przez tysiące lat działalność budowlana opierała się na metodach deterministycznych - precyzyjnych obliczeniach, detalach i ścisłej kontroli parametrów. W pierwszych wiekach naszej ery rzymscy inżynierowie stosowali zasady matematyczne do budowy akweduktów i mostów. W średniowieczu architekci dążyli do idealnych proporcji gotyckich katedr, a wraz z industrializacją XX wieku powstały systemy znormalizowanych norm i przepisów, które stały się podstawą budownictwa masowego.

Obecnie wektor rozwoju przesuwa się z poszukiwania związków stricte przyczynowo-skutkowych w kierunku analizy probabilistycznej, poszukiwania korelacji i ukrytych wzorców. Branża wchodzi w nową fazę - dane stają się kluczowym zasobem, a oparta na nich analityka zastępuje intuicyjne i lokalnie zoptymalizowane podejścia.



Rys. 10.1-1 Ukryty potencjał danych budowlanych: istniejące obliczenia w firmie to tylko wierzchołek góry lodowej dostępnej do analizy przez kierownictwo.

System informacyjny firmy jest jak góra lodowa (rys. 10.1-1): tylko niewielka część potencjału danych jest widoczna dla kierownictwa firmy, podczas gdy główna wartość jest ukryta w głębi. Ważne jest, aby oceniać dane nie tylko pod kątem ich bieżącego wykorzystania, ale także pod kątem możliwości, które uwalniają w przyszłości. To właśnie te firmy, które nauczą się wydobywać ukryte wzorce i tworzyć nowe spostrzeżenia z danych, stworzą trwałą przewagę konkurencyjną.

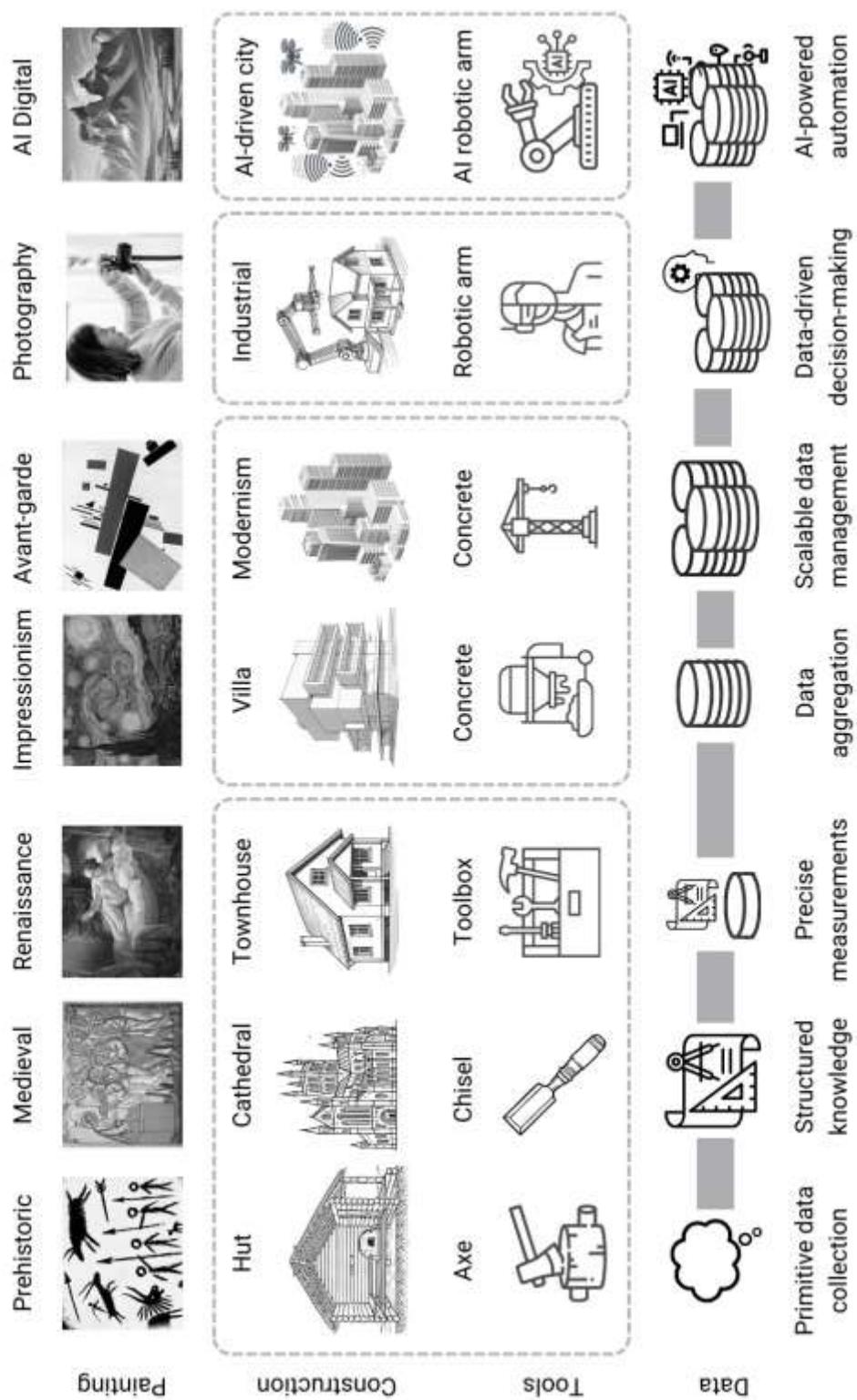
Znajdowanie ukrytych wzorców i nadawanie sensu danym to nie tylko praca z liczbami, ale proces twórczy, który wymaga abstrakcyjnego myślenia i umiejętności dostrzeżenia całego obrazu kryjącego się za różnymi elementami. W tym sensie ewolucję pracy z danymi można porównać do ewolucji sztuk wizualnych (Rys. 10.1-2).

Rozwój budownictwa jest niezwykle podobny do rozwoju sztuk pięknych. W obu przypadkach ludzkość przeszła od prymitywnych metod do zaawansowanych technologii wizualizacji i analizy. W czasach prehistorycznych ludzie używali rysunków jaskiniowych i prymitywnych narzędzi do rozwiązywania codziennych zadań. W okresie średniowiecza i renesansu poziom zaawansowania w architekturze i sztuce gwałtownie wzrósł. We wczesnym średniowieczu narzędzia budowlane ewoluowały od prostej siekierki do rozbudowanych zestawów narzędzi symbolizujących rozwój wiedzy technicznej.

Wiek realizmu był pierwszą rewolucją w sztukach wizualnych: artyści nauczyli się odtwarzać najdrobniejsze szczegóły, osiągając maksymalną wiarygodność. W budownictwie odpowiednikiem tego okresu były precyzyjne techniki inżynierijne, szczegółowe rysunki i ścisłe regulowane obliczenia, które stały się podstawą praktyki projektowej przez wieki.

Później impresjonizm zmienił samo postrzeganie rzeczywistości artystycznej: zamiast dosłownie renderować formę, artyści zaczęli uchwycić nastrój, światło i dynamikę, dążąc do odzwierciedlenia ogólnego wrażenia, a nie absolutnej dokładności. Podobnie, uczenie maszynowe w analityce budynków odchodzi od sztywnych modeli logicznych na rzecz rozpoznawania wzorców i wzorców probabilistycznych, które pozwalają "zobaczyć" ukryte zależności w danych, niedostępne w klasycznej analizie. Podejście to odzwierciedla idee minimalizmu i funkcjonalności Bauhausu, gdzie znaczenie (funkcja) jest ważniejsze niż forma. Bauhaus dążył do usunięcia tego, co zbędne, porzucenia ornamentyki na rzecz przejrzystości, utylitaryzmu i masowej atrakcyjności. Rzeczy musiały być zrozumiałe i użyteczne, bez nadmiaru - estetyka narodziła się z logiki projektowania i celu.

Wraz z pojawiением się fotografii pod koniec XIX wieku, sztuka zyskała nowe narzędzie do uchwycenia rzeczywistości z niespotykaną dotąd dokładnością i wywróciła do góry nogami podejście do sztuk wizualnych. Podobnie w budownictwie, rewolucja przemysłowa w XXI wieku prowadzi do wykorzystania technologii robotycznych, laserów, IoT, RFID i koncepcji takich jak Connected Construction, gdzie gromadzenie poszczególnych parametrów ewoluowało do skalowalnego inteligentnego przechwytywania pełnej rzeczywistości placu budowy.



Rys. 10.1-2 Era Ewolucji Sztuk Pięknych jest spójna z rozwojem podejścia do pracy z danymi w branży budowlanej.

Dziś, podobnie jak sztuki wizualne doświadczają ponownego przemyślenia wraz z pojawieniem się narzędzi AI i LLM, branża budowlana doświadcza kolejnego skoku kwantowego: inteligentne systemy napędzane sztuczną inteligencją (AI), czaty LLM umożliwiają przewidywanie, optymalizację i generowanie rozwiązań przy minimalnej interwencji człowieka.

Rola danych w projektowaniu i zarządzaniu zmieniała się radykalnie. Podczas gdy kiedyś wiedza była przekazywana ustnie i miała charakter empiryczny - tak jak rzeczywistość była uchwycona przez ręcznie malowane obrazy do XIX wieku - dziś nacisk kładziony jest na pełne cyfrowe uchwycenie "obrazu" konstrukcji. Z pomocą algorytmów uczenia maszynowego, ten cyfrowy obraz jest przekształcany w impresjonistyczną reprezentację rzeczywistości budowlanej - nie dokładną replikę, ale uogólnione, probabilistyczne zrozumienie procesów.

Szybko zbliżamy się do ery, w której procesy projektowania, konstruowania i obsługi budynków będą nie tylko rozszerzone, ale w dużej mierze napędzane przez systemy sztucznej inteligencji. Podobnie jak nowoczesna sztuka cyfrowa tworzona jest bez użycia pędzla - przy użyciu podpowiedzi tekstowych i modeli generatywnych - rozwiązania architektoniczne i inżynierijne przyszłości będą kształtowane przez kluczowe zapytania i parametry ustawiane przez użytkownika.

W XXI wieku dostęp do danych, ich interpretacja i jakość analityki stają się niezbędne dla powodzenia projektów. A o wartości danych nie decyduje ich ilość, ale zdolność specjalistów do ich analizy, weryfikacji i przełożenia na działania.

Podejście oparte na danych w budownictwie: nowy poziom infrastruktury

W historii ludzkości każdy taki skok technologiczny przynosił fundamentalne zmiany w gospodarce i społeczeństwie. Dziś jesteśmy świadkami nowej fali transformacji porównywalnej skalą do rewolucji przemysłowej z XIX wieku. O ile jednak sto lat temu głównym motorem zmian były siły mechaniczne i technologie energetyczne, o tyle obecnie są nim dane i sztuczna inteligencja.

Uczenie maszynowe, LLM i agenci AI zmieniają istotę aplikacji, sprawiając, że tradycyjne stopy oprogramowania (omówione w drugiej części książki) stają się zbędne (rys. 2.2-3). Cała logika danych jest skoncentrowana na agentach AI, a nie na zakodowanych na stałe regułach biznesowych (rys. 2.2-4).

W erze danych tradycyjne spojrzenie na aplikacje ulega zasadniczej transformacji. Zmierzamy w kierunku modelu, w którym nieporęczne, modułowe systemy korporacyjne nieuchronnie ustąpią miejsca otwartym, lekkim, niestandardowym rozwiązaniom.

W przyszłości pozostanie tylko podstawowa struktura danych, a cała interakcja z nią będzie odbywać się za pośrednictwem agentów pracujących bezpośrednio z bazą danych. Naprawdę wierzę, że cały stos aplikacji zniknie, ponieważ po prostu nie ma takiej potrzeby, gdy sztuczna inteligencja wchodzi w interakcję bezpośrednio z bazą danych. Spędziłem całą swoją karierę pracując w SaaS - budując firmy, pracując w nich i szczerze mówiąc, prawdopodobnie nie uruchomiłbym teraz nowego biznesu SaaS. I prawdopodobnie nie zainwestowałbym teraz w firmy SaaS. Sytuacja jest zbyt niepewna. Nie oznacza to, że w przyszłości nie będzie firm tworzących oprogramowanie, ale będą one wyglądać zupełnie inaczej. Przyszłe systemy będą bazami danych z logiką biznesową wprowadzoną do agentów [AI]. Agenci ci będą pracować z wieloma repozytoriami danych jednocześnie, nie ograniczając się do jednej bazy danych. Cała logika zostanie przeniesiona do warstwy AI [46].

- Matthew Berman, CEO Forward Future

Kluczową różnicą nowego paradygmatu jest minimalizacja balastu technologicznego. Zamiast monumentalnych, złożonych i zamkniętych systemów oprogramowania, otrzymamy elastyczne, otwarte i szybko konfigurowalne moduły, które dosłownie "żyją" wewnętrz przepływu danych (rys. 7.4-1 - Apache Airflow, NiFi). Architektura zarządzania procesami przyszłości przewiduje wykorzystanie mikroaplikacji - kompaktowych, specjalnie zbudowanych narzędzi, zasadniczo różniących się od masywnych i zamkniętych systemów ERP, PMIS, CDE, CAFM. Nowi agenci będą tak adaptacyjni, zintegrowani i specyficzni dla biznesu, jak to tylko możliwe (np. Low-Code/No-Code Rys. 7.4-6).

Cała logika biznesowa trafi do tych agentów [AI], a agenci ci będą wykonywać operacje CRUD [Create, Read, Update, and Delete] na wielu repozytoriach, co oznacza, że nie będą rozróżniać, który backend jest używany. Będą aktualizować wiele baz danych, a cała logika zakończy się na tak zwanym poziomie AI. A gdy warstwa AI znajdzie się tam, gdzie cała logika, ludzie zaczyną wymieniać backende. Już teraz obserwujemy dość wysoki odsetek wygranych rynkowych w backendach Dynamics i wykorzystaniu agentów, a my będziemy agresywnie podążać w tym kierunku, próbując połączyć to wszystko razem. Niezależnie od tego, czy chodzi o obsługę klienta, czy inne obszary, na przykład nie tylko CRM, ale także nasze rozwiązania finansowe i operacyjne. Ponieważ ludzie chcą więcej aplikacji biznesowych opartych na sztucznej inteligencji, w których warstwa logiczna może być napędzana przez sztuczną inteligencję i agentów sztucznej inteligencji. [...]. Jedną z najbardziej ekscytujących rzeczy dla mnie jest Excel z Pythonem, który jest porównywalny do GitHuba z Copilotem. To znaczy, co zrobiliśmy: teraz, gdy masz Excela, powinieneś go po prostu otworzyć, uruchomić Copilota i zacząć się nim bawić. Nie chodzi już tylko o zrozumienie dostępnych liczb - sam stworzy plan. Podobnie jak GitHub Copilot Workspace tworzy plan, a następnie go wykonuje, jest jak analityk danych, używający Excela jako narzędzia do wizualizacji wierszy i kolumn do analizy. Tak więc Copilot używa Excela jako narzędzia ze wszystkimi jego możliwościami, ponieważ może generować dane i ma interpreter Pythona.

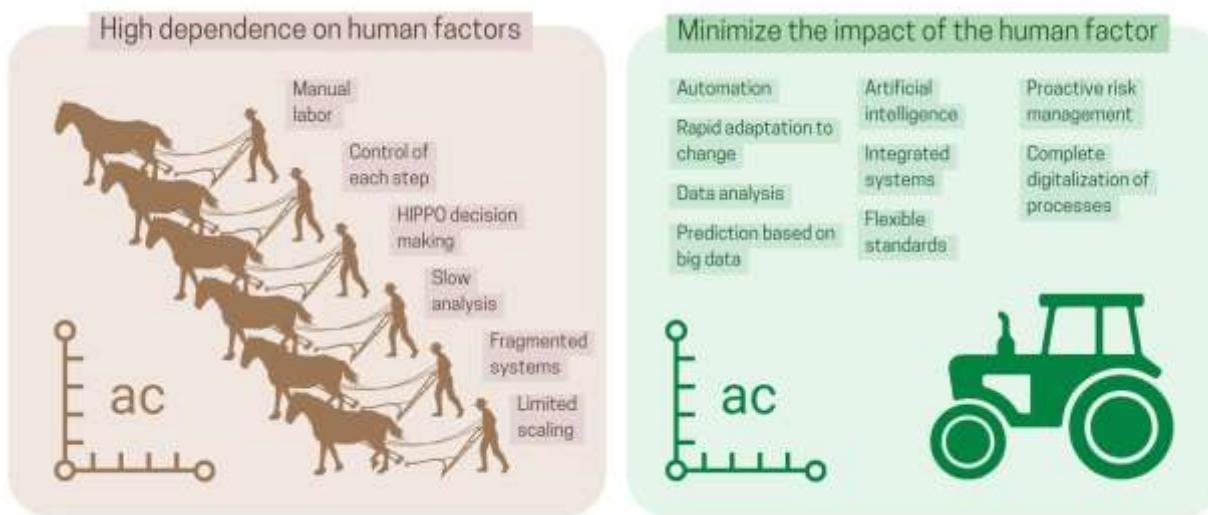
- Satya Nadella, CEO, Microsoft, wywiad z kanałem BG2, grudzień 2024. [28]

Transformacja, której jesteśmy świadkami w logice aplikacji biurowych - przejście od modułowych,

zamkniętych systemów do agentów AI pracujących bezpośrednio z otwartymi danymi - jest tylko częścią znacznie większego procesu. Nie chodzi tylko o zmianę interfejsów czy architektury oprogramowania: zmiany wpłyną na fundamentalne zasady organizacji pracy, podejmowania decyzji i zarządzania biznesem. W budownictwie doprowadzi to do logiki opartej na danych, w której dane staną się centralnym elementem procesów od projektowania po zarządzanie zasobami i monitorowanie budowy.

Cyfrowe biuro nowej generacji: jak sztuczna inteligencja zmienia przestrzeń roboczą

Prawie sto lat temu ludzkość doświadczała już podobnej rewolucji technologicznej. Przejście od silników parowych do elektrycznych zajęło ponad cztery dekady, ale ostatecznie stało się katalizatorem bezprecedensowego wzrostu produktywności - głównie dzięki decentralizacji mocy energetycznej i elastyczności nowych rozwiązań. Zmiana ta nie tylko zmieniła bieg historii, przenosząc większość populacji z obszarów wiejskich do miejskich, ale także położyła podwaliny pod nowoczesną gospodarkę. Historia technologii to podróż od pracy fizycznej do automatyzacji i inteligentnych systemów. Tak jak traktor zastąpił wiele maszyn uprawiających ziemię, tak nowoczesna technologia cyfrowa zastępuje tradycyjne, biurowe metody zarządzania budową (Rys. 10.1-3). Jeszcze na początku XX wieku większość światowej populacji uprawiała ziemię ręcznie, aż do momentu, gdy w latach 30. ubiegłego wieku rozpoczęto mechanizację pracy za pomocą maszyn i ciągników.



Rys. 10.1-3 Tak jak traktor zastąpił dziesiątki ludzi na początku XX wieku, tak uczenie maszynowe zastąpi tradycyjne metody zarządzania biznesem i projektami w XXI wieku.

Tak jak sto lat temu ludzkość przeszła od uprawy pojedynczych działek za pomocą prymitywnych narzędzi do rolnictwa na dużą skalę z wykorzystaniem maszyn, tak dziś przechodzimy od przetwarzania rozproszonych "silosów" informacji do pracy z tablicami danych za pomocą potężnych "traktorów" - ETL-pipeline i algorytmów sztucznej inteligencji.

Jesteśmy u progu podobnego skoku - ale na płaszczyźnie cyfrowej: od tradycyjnego, ręcznego zarządzania biznesem do modeli opartych na danych.

Droga do pełnej architektury opartej na danych będzie wymagała czasu, inwestycji i wysiłków organizacyjnych. Droga ta otwiera jednak drogę nie tylko do stopniowej poprawy, ale do jakościowego skoku w kierunku większej wydajności, przejrzyści i łatwości zarządzania procesami budowlanymi. Wszystko to pod warunkiem systematycznego wprowadzania narzędzi cyfrowych i rezygnacji z przestarzałych praktyk biznesowych.

Parametryzacja zadań, ETL, LLM, komponenty IoT, RFID, tokenizacja, big **data** i uczenie maszynowe przekształca tradycyjne budownictwo w **oparte na danych budownictwo**, w którym każdy szczegół projektu i działalności budowlanej będzie kontrolowany i optymalizowany przez dane.

Kiedyś analiza informacji zajmowała tysiące roboczogodzin. Obecnie zadania te są wykonywane przez algorytmy i LLM, które za pomocą podpowiedzi przekształcają różne zbiory danych w strategiczne źródła. W świecie technologii dzieje się to samo, co w rolnictwie: przechodzimy od motyki do zautomatyzowanego agrobiznesu. Tak samo praca biurowa w budownictwie - od plików Excel i ręcznych podsumowań - przechodzi do inteligentnego systemu, w którym dane są gromadzone, czyszczone, strukturyzowane i przekształcane w spostrzeżenia.

Już dziś firmy powinny zacząć "uprawiać" pola informacyjne poprzez gromadzenie danych wysokiej jakości i strukturyzację informacji, "nawozić" je narzędziami do czyszczenia i normalizacji, a następnie "zbierać" je w postaci analityki predykcyjnej i zautomatyzowanych rozwiązań. Jeśli nowoczesny rolnik z maszyną jest w stanie zastąpić sto glebogryzarek, inteligentne algorytmy będą w stanie zdjąć rutynę z pracowników i przenieść ich do roli strategicznych menedżerów przepływów informacji.

Ważne jest jednak, aby zrozumieć, że stworzenie prawdziwie opartej na danych organizacji nie jest szybkim procesem. Jest to długoterminowy kierunek strategiczny, podobny do tworzenia nowego miejsca do uprawy nowego lasu (rys. 1.2-5) systemów, gdzie każde "drzewo" w tym ekosystemie jest oddzielnym procesem, kompetencją lub narzędziem, którego wzrost i rozwój wymaga czasu. Podobnie jak w przypadku prawdziwego lasu, sukces zależy nie tylko od jakości materiału do sadzenia (technologia), ale także od gleby (kultura korporacyjna), klimatu (środowisko biznesowe) i opieki (podejście systemowe).

Firmy nie będą już mogły polegać wyłącznie na zamkniętych, gotowych rozwiązaniach. W przeciwieństwie do poprzednich etapów rozwoju technologicznego, obecne przejście - do otwartego dostępu do danych, sztucznej inteligencji i Open Source - raczej nie będzie wspierane przez dużych dostawców, ponieważ bezpośrednio zagraża ich ugruntowanym modelom biznesowym i podstawowym strumieniom przychodów.

Jak wynika z badania Harvard Business School [40], które zostało już omówione w rozdziale poświęconym czwartej i piątej rewolucji technologicznej, koszt stworzenia od podstaw najczęściej wykorzystywanych rozwiązań Open Source dla wszystkich firm wyniósłby około 4,15 miliarda dolarów. Jeśli jednak wyobrażymy sobie, że każda firma opracowałaby własne alternatywy bez dostępu do istniejących narzędzi Open Source, co działo się przez ostatnie dziesięciolecia, całkowite koszty biznesowe mogłyby osiągnąć kolosalną kwotę 8,8 biliona dolarów - jest to cena irracjonalnego popytu, na jaką można wycenić rynek oprogramowania.

Postęp technologiczny nieuchronnie doprowadzi do ponownego przemyślenia ustalonych modeli biznesowych. Podczas gdy wcześniej firmy były w stanie zarabiać na złożonych, nieprzejrzystych procesach i zamkniętych danych, wraz z rozwojem sztucznej inteligencji i analityki takie podejście staje się coraz mniej opłacalne.

W wyniku demokratyzacji dostępu do danych i narzędzi, tradycyjny rynek sprzedaży oprogramowania może się znacznie skurczyć. W tym samym czasie rozwijać się będzie jednak nowy rynek - rynek ekspertyz cyfrowych, dostosowywania, integracji i projektowania rozwiązań. W tym przypadku wartość nie będzie pochodzić ze sprzedaży licencji, ale z możliwości budowania elastycznych, otwartych i adaptowalnych procesów cyfrowych. Podobnie jak elektryfikacja i pojawienie się traktorów stworzyły nowe gałęzie przemysłu, tak samo zastosowanie big data, AI i LLM otworzy zupełnie nowe horyzonty dla biznesu w branży budowlanej, co będzie wymagało nie tylko inwestycji technologicznych, ale także głębskiej transformacji sposobu myślenia, procesów i struktur organizacyjnych. Firmy i specjaliści, którzy zdadzą sobie z tego sprawę i zaczną działać już dziś, będą liderami jutra.

W świecie, w którym otwarte dane stają się głównym atutem, dostępność informacji zmieni zasady gry. Inwestorzy, klienci i organy regulacyjne będą coraz bardziej domagać się przejrzystości, a algorytmy uczenia maszynowego będą w stanie automatycznie identyfikować rozbieżności w szacunkach, harmonogramach i kosztach. Stwarza to warunki dla nowego etapu transformacji cyfrowej, która stopniowo prowadzi nas do "uberyzacji" branży budowlanej.

Otwarte dane i Uberyzacja stanowią zagrożenie dla istniejącej działalności budowlanej

Budownictwo staje się procesem zarządzania informacją. Im większa dokładność, jakość i kompletność danych, tym sprawniejsze projektowanie, obliczenia, kosztorysy, wznoszenie i eksploatacja budynków. W przyszłości kluczowym zasobem nie będzie dźwig, beton i pręty zbrojeniowe, ale zdolność do gromadzenia, analizowania i wykorzystywania informacji.

W przyszłości klienci firm budowlanych - inwestorzy i klienci finansujący budowę - nieuchronnie wykorzystają wartość otwartych danych i analityki danych historycznych. Otworzy to możliwości automatyzacji obliczania harmonogramów i kosztów projektów, bez angażowania firm budowlanych w kwestie kosztorysowania, co pomoże kontrolować koszty i szybciej identyfikować zbędne koszty.

Wyobraźmy sobie plac budowy, na którym skanery laserowe, quadrocoptery i systemy fotogrametryczne zbierają w czasie rzeczywistym dokładne dane na temat ilości użytego betonu. Informacje te są automatycznie przekształcane w proste płaskie modele MESH z metadanymi, omijając uciążliwe systemy CAD (BIM), bez zależności od złożonych jąder geometrycznych, ERP lub PMIS. Dane zebrane z placu budowy są centralnie przesyłane do jednego ustrukturyzowanego repozytorium, dostępnego dla klienta do niezależnej analizy, gdzie przesyłane są rzeczywiste ceny z różnych sklepów budowlanych, z parametrami od stopy finansowania kredytowego po dynamicznie zmieniające się czynniki, takie jak warunki pogodowe, notowania giełdowe materiałów budowlanych, taryfy logistyczne i statystyczne sezonowe wahania cen pracy. W takich warunkach wszelkie rozbieżności między projektem a rzeczywistymi ilościami materiałów stają się natychmiast oczywiste, uniemożliwiając manipulowanie szacunkami zarówno na etapie projektowania, jak i po przekazaniu projektu. W rezultacie przejrzystość procesu budowlanego osiąga się nie poprzez armię nadzorców i menedżerów, ale poprzez obiektywne dane cyfrowe, które zminimalizują

czynnik ludzki i możliwość spekulacji.

W przyszłości tego rodzaju prace związane z kontrolą danych będą raczej wykonywane przez menedżerów danych po stronie klienta (rys. 1.2-4 Menedżer CQMS). Dotyczy to zwłaszcza kalkulacji i szacunków projektów: tam, gdzie kiedyś istniał cały dział estymatorów, jutro będą już narzędzia do uczenia maszynowego i prognozowania, które ustalą limity cenowe dla firm budowlanych.

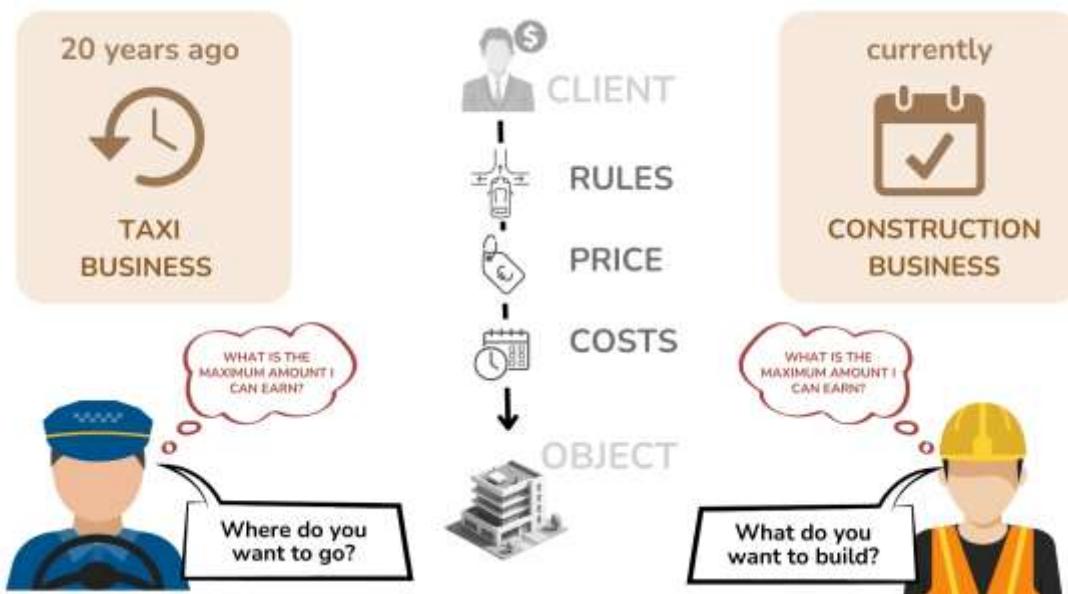
Biorąc pod uwagę rozdrobniony charakter branży [budowlanej], w której większość systemów i podsystemów jest dostarczana przez MŚP, strategia cyfrowa musi pochodzić od klienta. Klienci muszą stworzyć warunki i mechanizmy, aby odblokować cyfrowe możliwości łańcucha dostaw [20].

- Andrew Davis i Giuliano Denicol, Accenture "Tworzenie większej wartości poprzez projekty kapitałowe"

Taka otwartość i przejrzystość danych stanowi zagrożenie dla firm budowlanych, które są przyzwyczajone do zarabiania pieniędzy na nieprzejrzystych procesach i mylących raportach, w których spekulacje i ukryte koszty mogą być ukryte za złożonymi i zamkniętymi formatami oraz modułowymi, zastrzeżonymi platformami danych. Dlatego też firmy budowlane, podobnie jak sprzedawcy promujący rozwiązania Open Source, raczej nie będą zainteresowane pełnym wdrożeniem otwartych danych do swoich procesów biznesowych. Jeśli dane są dostępne i łatwe do przetworzenia dla klienta, mogą być sprawdzane automatycznie, co wyeliminuje możliwość przeszacowania wolumenów i manipulowania szacunkami.

Według raportu Światowego Forum Ekonomicznego "Shaping the Future of Construction" (2016) [5], jednym z kluczowych wyzwań dla branży pozostaje pasywna rola klienta. Niemniej jednak, to klienci powinni wziąć większą odpowiedzialność za wyniki projektów - od wczesnego planowania, przez wybór zrównoważonych modeli interakcji, po monitorowanie wydajności. Bez aktywnego udziału właścicieli projektów, systemowa transformacja branży budowlanej jest niemożliwa.

Utrata kontroli nad wolumenem i kalkulacją kosztów już przekształciła inne branże w ciągu ostatnich 20 lat, umożliwiając klientom bezpośrednie, bez pośredników, utrudnianie realizacji ich celów. Cyfryzacja i przejrzystość danych przekształciły wiele tradycyjnych modeli biznesowych, takich jak taksówkarze wraz z pojawiением się Ubera (Rys. 10.1-4), hotelarze wraz z pojawiением się Airbnb oraz detaliści i sklepy wraz z pojawiением się Amazona, a także banki wraz z pojawiением się neobanków i zdecentralizowanych ekosystemów fintech, w których bezpośredni dostęp do informacji oraz automatyzacja kalkulacji czasu i kosztów znacznie zmniejszyły rolę pośredników.



Rys. 10.1-4 Branża budowlana stanie w obliczu uberyzacji, z którą 10 lat temu musieli zmierzyć się taksówkarze, hotelarze i sprzedawcy detaliczni.

Proces demokratyzacji dostępu do danych i narzędzi do ich przetwarzania jest nieunikniony, a z czasem otwarte dane dotyczące wszystkich komponentów projektu staną się wymogiem klientów i nowym standardem. Dlatego kwestie wprowadzania otwartych formatów i przejrzystych obliczeń będą promowane przez inwestorów, klientów, banki i fundusze private equity (private equity) - tych, którzy są końcowymi użytkownikami budowanych obiektów, a następnie eksploatują obiekt przez dziesięciolecia.

Główni inwestorzy, klienci i banki już teraz domagają się przejrzystości w branży budowlanej. Według badania Accenture "Tworzenie większej wartości poprzez projekty kapitałowe" (2020) [20], przejrzyste i wiarygodne dane stają się decydującym czynnikiem przy podejmowaniu decyzji inwestycyjnych w budownictwie. Jak podkreślają eksperci, wiarygodne i skuteczne zarządzanie projektami nie jest możliwe bez przejrzystości, zwłaszcza w czasach kryzysu. Ponadto właściciele aktywów i wykonawcy coraz częściej dążą do zawierania umów, które zachęcają do udostępniania danych i współpracy analitycznej, odzwierciedlając rosnące wymagania ze strony inwestorów, banków i organów regulacyjnych w zakresie odpowiedzialności i przejrzystości.

Przemieszczanie się inwestora, klienta od pomysłu do gotowego budynku, w przyszłości będzie przypominało podróżowanie na autopilote - bez kierowcy w postaci firmy budowlanej, zapowiada się uniezależnienie od spekulacji i niepewności.

Era otwartych danych i automatyzacji nieuchronnie zmieni branżę budowlaną, podobnie jak miało to już miejsce w bankowości, handlu, rolnictwie i logistyce. W tych branżach rola pośredników i tradycyjne sposoby prowadzenia działalności ustępują miejsca automatyzacji i robotyzacji, nie pozostawiając miejsca na nieuzasadnione marże i spekulacje.

Dane i procesy we wszystkich ludzkich działaniach gospodarczych nie różnią się od tego, z czym mają do czynienia profesjonalisci z branżą budowlaną. W dłuższej perspektywie firmy budowlane, które dziś

dominują na rynku poprzez ustalanie cen i standardów jakości usług, mogą stracić swoją rolę kluczowego pośrednika między klientem a jego projektem budowlanym.

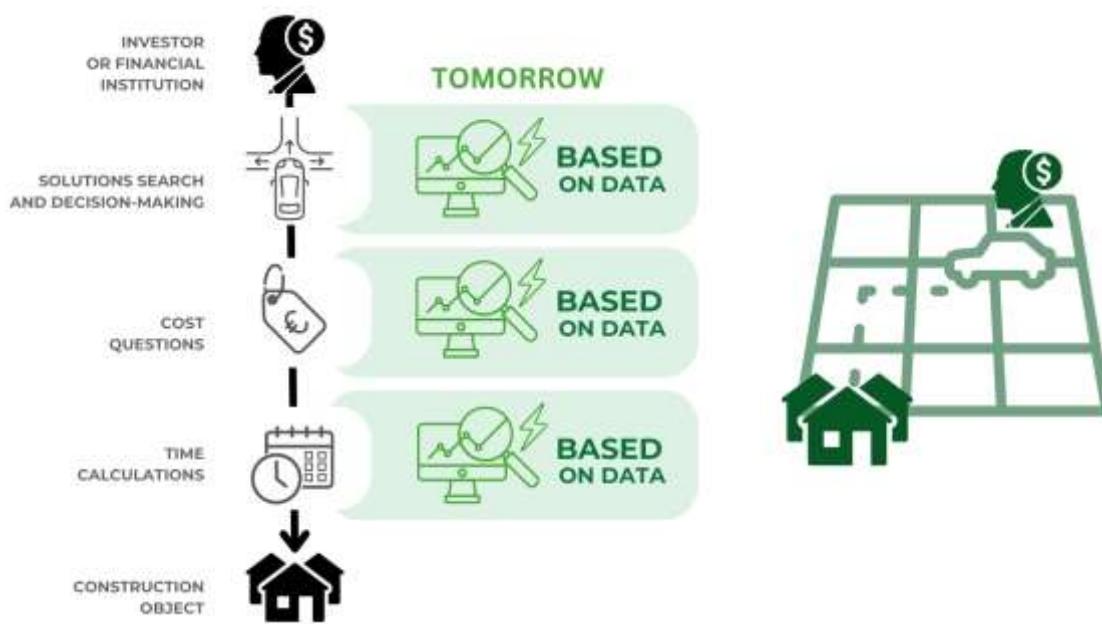
Nierozwiązane problemy uberyzacji ostatnią szansą na wykorzystanie czasu na transformację

Wróćmy jednak do realiów branży budowlanej. Podczas gdy samojezdne samochody, zdecentralizowane systemy finansowe i rozwiązania oparte na sztucznej inteligencji pojawiają się w niektórych sektorach gospodarki, znaczna część firm budowlanych nadal pozostaje organizacjami opartymi na papierze, w których kluczowe decyzje podejmowane są bardziej w oparciu o intuicję i doświadczenie poszczególnych specjalistów.

W tym paradygmacie nowoczesną firmę budowlaną można porównać do 20-letniej firmy taksówkarskiej, która kontroluje zasoby, trasy i czasy dostaw oraz jest odpowiedzialna za czas i koszt "podróży" - od pomysłu na projekt (proces logistyczny i instalacyjny) do dostarczenia projektu. Podobnie jak GPS (w budownictwie IoT, RFID) i algorytmy uczenia maszynowego w obliczeniach czasu/kosztów przekształciły kiedyś dziedzinę transportu, dane, algorytmy i sztuczna inteligencja -agenci mają potencjał, aby przekształcić zarządzanie budową - od intuicyjnych ocen do predykcyjnych, kierowanych modeli. W ciągu ostatnich 20 lat w wielu branżach - finansach, rolnictwie, handlu detalicznym i logistyce - stopniowo zanikała możliwość spekulacji dzięki nieprzejrzystości danych. Ceny, koszty dostawy lub transakcje finansowe są obliczane automatycznie i statystycznie - w ciągu zaledwie kilku sekund na platformach cyfrowych.

Patrząc w przyszłość, firmy budowlane muszą zdać sobie sprawę, że demokratyzacja dostępu do danych i narzędzi do ich analizy zakłóci tradycyjne podejście do szacowania kosztów i harmonogramów projektów oraz wyeliminuje możliwość spekulacji na nieprzejrzystych danych dotyczących wolumenu i cen.

Podobnie jak jazda po regulowanej drodze bez interwencji kierowcy, procesy budowlane przyszłości będą coraz bardziej przypominać system "Uberised" - ze zautomatyzowanym szacowaniem czasu i kosztów, przejrzystym trasowaniem zadań i minimalną zależnością od czynnika ludzkiego. Zmieni to sam charakter "podróży" od pomysłu do realizacji - czyniąc ją bardziej przewidywalną, zarządzalną i opartą na danych.



Rys. 10.1-5 Koszt i czas "podróży" podczas budowy zostaną określone przy użyciu uczenia maszynowego i narzędzi statystycznych.

Wraz ze stopniowym wprowadzaniem nowych regulacji i wymogów w niemal każdym kraju na świecie, które obligują do przekazywania klientom lub bankom finansującym projekty budowlane modeli CAD- (BIM), klient i zleceniodawca mają możliwość samodzielnego zapewnienia przejrzystości kalkulacji kosztów i zakresu. Jest to szczególnie ważne dla dużych klientów i inwestorów, którzy posiadają wystarczające kompetencje i narzędzia do szybkiej analizy wolumenów i monitorowania cen rynkowych. Dla firm realizujących standardowe projekty na dużą skalę - sklepy, biurowce, osiedla mieszkaniowe - takie praktyki stają się standardem.

W miarę jak zawartość informacyjna modeli staje się coraz bardziej kompletna i ustandaryzowana, możliwość manipulacji i spekulacji praktycznie zanika. Transformacja cyfrowa stopniowo zmienia zasady gry w branży budowlanej, a firmy, które nie dostosują się do tych zmian, mogą stanąć przed poważnymi wyzwaniami.

Rosnąca konkurencja, zakłócenia technologiczne i kurczące się marże mogą potencjalnie wpływać na zrównoważony rozwój biznesu. Przy ograniczonej płynności finansowej coraz więcej graczy w branży zwraca się w stronę automatyzacji, analityki i technologii danych jako sposobu na poprawę wydajności i przejrzystości procesów. Narzędzia te stają się ważnym zasobem pozwalającym zachować konkurencyjność w zmieniającym się środowisku gospodarczym.

Być może nie warto czekać, aż okoliczności zewnętrzne wymuszą pilne działania - znacznie skuteczniej jest zacząć przygotowania już dziś, wzmacniając kompetencje cyfrowe, wdrażając nowoczesne rozwiązania i budując kulturę skoncentrowaną na danych.

Jedną z ostatnich kluczowych barier technologicznych dla cyfrowej transformacji branży budowlanej na dużą skalę, która dotknie każdą firmę w nadchodzących latach, jest problem automatycznej klasyfikacji elementów projektów budowlanych.

Bez niezawodnej, dokładnej i skalowalnej klasyfikacji niemożliwe jest stworzenie podstaw do pełnoprawnej analityki, automatyzacji procesów i zarządzania cyklem życia przy użyciu sztucznej inteligencji i modeli predykcyjnych. Dopóki klasyfikacja obiektów wciąż zależy od ręcznej interpretacji dokonywanej przez doświadczonych specjalistów - brygadistów, projektantów, kosztorysantów - branża budowlana wciąż ma szansę. Czas ten można wykorzystać na przygotowanie się do nieuniknionych zmian: rosnących wymagań w zakresie przejrzystości, demokratyzacji narzędzi i danych oraz pojawienia się automatycznych systemów klasyfikacji, które radykalnie zmienią zasady gry.

Zadanie automatycznej klasyfikacji elementów świata budowlanego jest porównywalne pod względem złożoności z rozpoznawaniem obiektów w bezzałogowych systemach jezdnych, co stanowi jedno z głównych wyzwań. Wyobraźmy sobie bezzałogowy samochód podróżujący z punktu A do punktu B (rys. 10.1-5). Obecne systemy automatycznej jazdy mają problem z klasyfikacją obiektów rozpoznawanych przez lidary i kamery. Nie wystarczy, że samochód po prostu "zobaczy" przeszkodę lub punkt orientacyjny; musi bezbłędnie wiedzieć, czy przed nim znajduje się pieszy, znak drogowy czy kosz na śmieci.

Podobne fundamentalne wyzwanie stoi przed całą branżą budowlaną. Elementy projektu - takie jak okna, drzwi czy kolumny - mogą być rejestrowane w dokumentacji, reprezentowane w modelach CAD, fotografowane na placu budowy lub rozpoznawane w chmurach punktów ze skanowania laserowego. Jednak ich wizualne lub przybliżone rozpoznanie geometryczne nie wystarcza do zbudowania prawdziwie zautomatyzowanego systemu zarządzania projektami. Konieczne jest zapewnienie, że każdy element jest dokładnie i konsekwentnie kategoryzowany do typu, który będzie jednoznacznie identyfikowalny we wszystkich kolejnych procesach - od szacunków i specyfikacji po logistykę, kontrolę zapasów i, co najważniejsze, operacje (rys. 4.2-6).

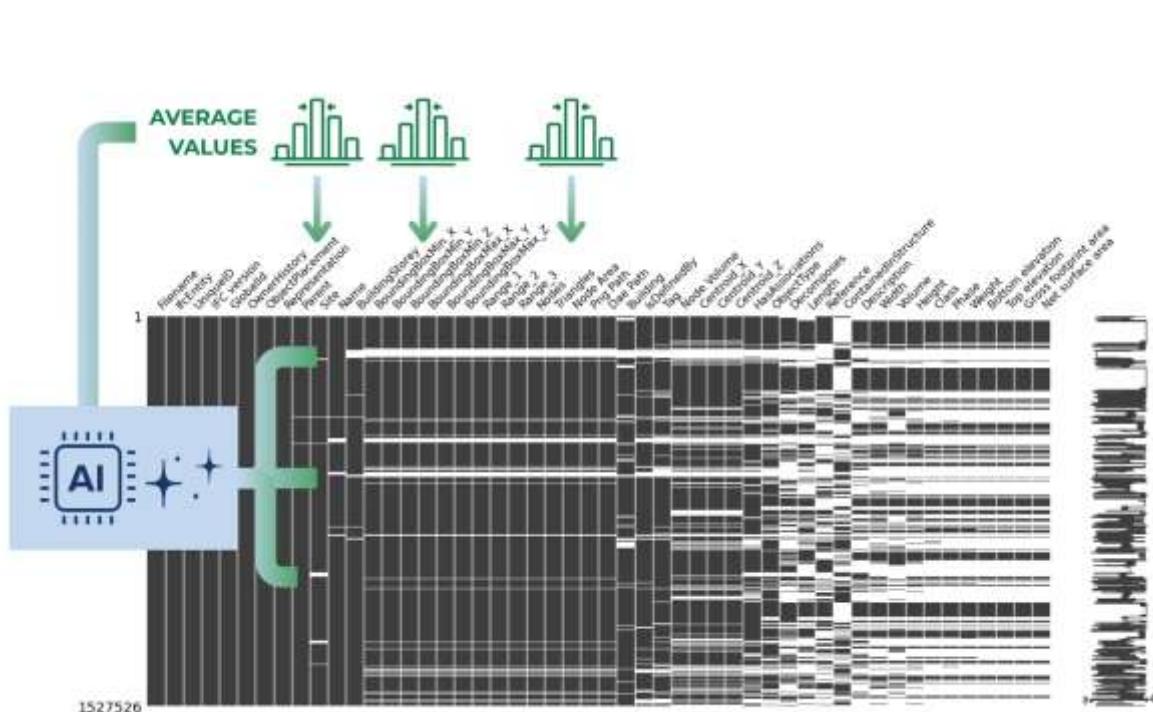
To właśnie na tym etapie - przejścia od rozpoznawania do znaczącej klasyfikacji - pojawia się jedna z kluczowych przeszkód. Nawet jeśli systemy cyfrowe są technicznie zdolne do rozróżniania i identyfikowania obiektów w modelach i na placu budowy, główna trudność polega na prawidłowej i stabilnej kontekstowej definicji typu elementu dla różnych środowisk oprogramowania.. Na przykład, drzwi mogą być oznaczone przez projektanta w modelu CAD jako element kategorii "drzwi", ale po przeniesieniu do systemu ERP lub PMIS mogą być błędnie wpisane, albo z powodu błędu projektanta, albo z powodu niespójności między systemami. Co więcej, element często traci niektóre ważne atrybuty lub zniką z systemu księgowego podczas eksportu i importu danych. Prowadzi to do luki w przepływie danych i podważa zasadę kompleksowej cyfryzacji procesów budowlanych. Tworzy to krytyczną lukę między "widocznym" a "zrozumiałym" znaczeniem semantycznym, co podważa integralność danych i znacznie komplikuje automatyzację procesów w całym cyklu życia projektu budowlanego.

Rozwiązanie wyzwania związanego z uniwersalną klasyfikacją elementów budowlanych przy użyciu technologii big data i uczenia maszynowego (rys. 10.1-6) będzie katalizatorem transformacji w całej branży - i być może nieoczekiwany odkryciem dla wielu firm budowlanych. Ujednolicony, możliwy do nauczenia się system klasyfikacji będzie podstawą skalowalnej analityki, cyfrowego zarządzania i przyjęcia sztucznej inteligencji do codziennych praktyk organizacji budowlanych.

NVIDIA i inni liderzy technologiczni już dostarczają rozwiązania w innych branżach, które mogą automatycznie kategoryzować i strukturyzować ogromne ilości informacji tekstowych i wizualnych.

Przykładowo, model NeMo Curator firmy NVIDIA [161] specjalizuje się w automatycznej klasyfikacji i kategoryzacji danych na predefiniowane kategorie, odgrywając kluczową rolę w optymalizacji potoków przetwarzania informacji w celu dostrajania i wstępnego uczenia generatywnych modeli sztucznej inteligencji. Platforma Cosmos jest szkolona na rzeczywistych scenach wideo i 3D [162], zapewniając podstawę dla autonomicznych systemów i cyfrowych bliźniaków, które są już budowane w ekosystemie NVIDIA. NVIDIA Omniverse, która do 2025 r. stała się wiodącym narzędziem do pracy z formatem USD, uniwersalnym opisem sceny, który może ostatecznie zastąpić format IFC w procesach przesyłania informacji projektowych. Wraz z Isaac Sim, zrobotyzowanym symulatorem procesów [163], rozwiązania takie jak NeMo Curator, Cosmos i Omniverse reprezentują nowy poziom automatyzacji: od czyszczenia i filtrowania danych po generowanie zestawów szkoleniowych, modelowanie właściwości obiektów i szkolenie robotów na placu budowy. Wszystkie te narzędzia są bezpłatne i otwarte, co znacznie zmniejsza bariery w ich stosowaniu w praktyce inżynierowej i budowlanej.

Automatyczna klasyfikacja danych na poziomie ustrukturyzowanych tabel nie jest tak trudna, jak mogłoby się wydawać na pierwszy rzut oka. Jak pokazaliśmy w poprzednim rozdziale (rys. 9.1-10), możliwe jest uzupełnienie brakujących lub nieprawidłowych wartości klas na podstawie podobnych parametrów innych elementów, jeśli istnieją zgromadzone dane historyczne. Jeśli elementy o podobnej charakterystyce zostały już poprawnie sklasyfikowane w kilku zakończonych projektach, system może z dużym prawdopodobieństwem zasugerować odpowiednią wartość dla nowego lub niekompletnego elementu (rys. 10.1-6). Taka logika, oparta na uśrednionych wartościach i analizie kontekstu, może być szczególnie skuteczna podczas masowego przetwarzania danych tabelarycznych pochodzących z szacunków, specyfikacji lub modeli CAD.

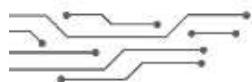


Rys. 10.1-6 Uczenie maszynowe może pomóc w automatycznym znalezieniu średnich wartości dla niewypełnionych (białe pola) parametrów tabeli na podstawie poprzednich projektów.

W kontekście tak szybkiego postępu w uczeniu maszynowym, jasne jest, że w 2025 roku naiwnością jest wierzyć, że problem automatycznej klasyfikacji elementów budynków pozostanie nierozerwiany przez długi czas. Tak, obecne algorytmy nie osiągnęły jeszcze pełnej dojrzałości, zwłaszcza w przypadku niekompletnych lub heterogenicznych danych, ale okno możliwości adaptacji szybko się zamknie.

Firmy, które już inwestują w gromadzenie, czyszczenie i organizowanie swoich danych, a także wdrażają narzędzia do automatyzacji ETL, będą miały przewagę. Reszta ryzykuje pozostać w tyle - tak jak kiedyś firmy nie sprostały wyzwaniom cyfrowej transformacji w branży transportowej i finansowej.

Ci, którzy nadal polegają na ręcznym zarządzaniu danymi i tradycyjnych metodach szacowania kosztów i czasu, ryzykują, że znajdą się w sytuacji flot taksówkowych z 2000 roku, niezdolnych do przystosowania się do ery aplikacji mobilnych i zautomatyzowanych obliczeń tras do początku 2020 roku.



ROZDZIAŁ 9.2.

PRAKTYCZNY PRZEWODNIK PO WDRAŻANIU PODEJŚCIA OPARTEGO NA DANYCH

Od teorii do praktyki: mapa drogowa cyfrowej transformacji w budownictwie

Branża budowlana stopniowo wkracza w nową fazę rozwoju, w której znane procesy są coraz częściej uzupełniane - a czasem nawet zastępowane - przez platformy cyfrowe i przejrzyste modele interakcji. Stawia to przed firmami nie tylko wyzwania, ale także znaczące możliwości. Organizacje, które już dziś budują długoterminową strategię cyfrową, będą w stanie nie tylko utrzymać swoją pozycję na rynku, ale także ją poszerzyć, oferując klientom nowoczesne podejście i niezawodne, wspierane technologicznie rozwiązania.

Ważne jest, aby zdać sobie sprawę, że znajomość koncepcji i technologii jest tylko punktem wyjścia. Menedżerowie i specjalisi stają przed praktycznym pytaniem: od czego zacząć wdrażanie i jak przekształcić teoretyczne pomysły w rzeczywistą wartość. Ponadto coraz częściej pojawia się pytanie: na czym będzie opierać się biznes, jeśli tradycyjne metody kalkulacji kosztów i harmonogramów mogą zostać w każdej chwili zmienione przez klienta.

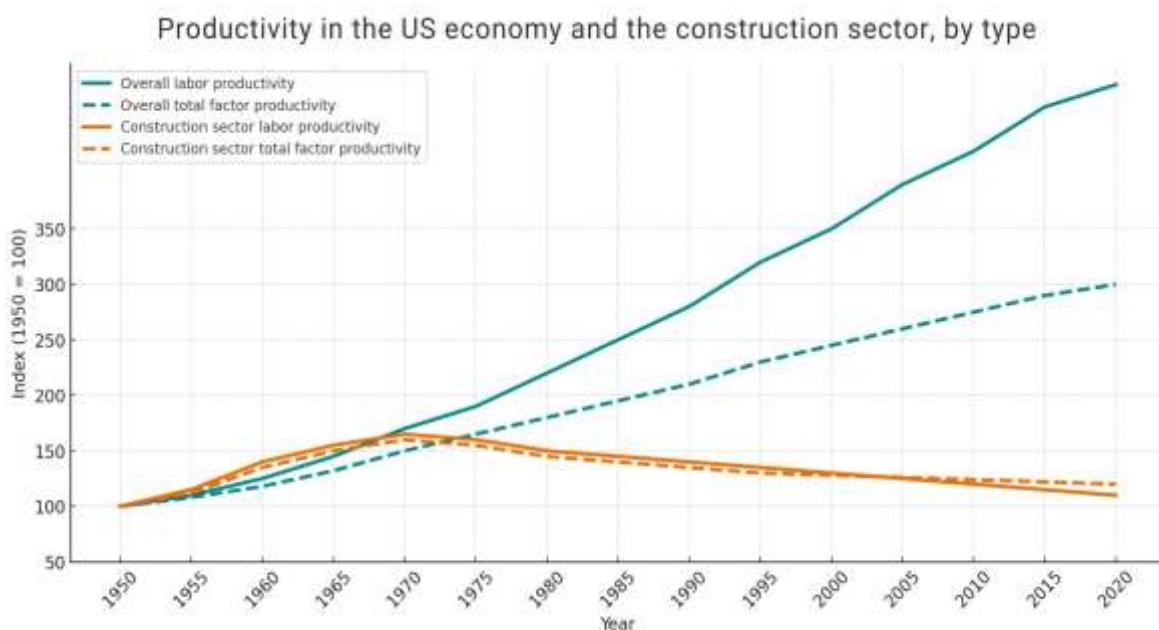
Odpowiedź prawdopodobnie leży nie tyle w technologii, co w tworzeniu nowej kultury zawodowej, w której praca z danymi jest postrzegana jako integralna część codziennej praktyki. To właśnie brak uwagi poświęcanej technologiom cyfrowym i innowacjom zaszczepił w branży budowlanej poważne zacofanie, które obserwowano w ostatnich dziesięcioleciach [43].

Według McKinsey, wydatki na badania i rozwój w branży budowlanej wynoszą mniej niż 1% przychodów, podczas gdy w przemyśle motoryzacyjnym i lotniczym wskaźnik ten sięga 3,5-4,5%. Podobnie, koszty IT w budownictwie pozostają na poziomie poniżej 1% całkowitych przychodów [107].

W rezultacie spada nie tylko poziom automatyzacji, ale także wydajność pracy w budownictwie, a do 2020 r. pracownik budowlany będzie produkował mniej niż pół wieku temu (rys. 10.2-1)

Takie problemy z produktywnością w sektorze budowlanym są wspólne dla większości krajów rozwiniętych i rozwijających się (produktywność w budownictwie spadła w 16 z 29 krajów OECD (Rys. 2.2-1)) i wskazują nie tylko na brak technologii, ale także na potrzebę zmian systemowych w samym podejściu do zarządzania, szkoleń i innowacji.

Sukces transformacji cyfrowej zależy nie tyle od liczby i dostępności narzędzi, ale od zdolności organizacji do przeglądu swoich procesów i rozwijania kultury otwartej na zmiany. To nie sama technologia jest kluczowa, ale ludzie i procesy, które zapewniają jej efektywne wykorzystanie, wspierają ciągłe uczenie się i zachęcają do akceptacji nowych pomysłów.



Rys. 10.2-1 Paradoks wydajności pracy i całkowitej produktywności zasobów w amerykańskiej gospodarce i sektorze budowlanym (1950-2020) (na podstawie [43]).

We wczesnych częściach książki model środowiska biznesowego został porównany do ekosystemu leśnego (Rys. 2.1-2, Rys. 1.2-4, Rys. 1.3-2). W zdrowym lesie okresowe pożary, pomimo całej swojej niszczycielskiej mocy, odgrywają kluczową rolę w długoterminowej odnowie. Oczyszczają glebę ze starej roślinności, przywracając zmagazynowane składniki odżywcze i tworzą przestrzeń dla nowego życia. Niektóre gatunki roślin ewoluowały nawet tak, że ich nasiona otwierają się tylko wtedy, gdy są narażone na wysokie temperatury pożaru - naturalny mechanizm, który zapewnia idealny czas na kiełkowanie.

Podobnie w biznesie, kryzysy mogą działać jak "kontrolowane wypalenie", zachęcając do pojawiania się nowych podejść i firm, które nie są związane z przestarzałymi systemami. Takie okresy zmuszają do porzucenia nieefektywnych praktyk, uwalniając zasoby na innowacje. Tak jak las po pożarze zaczyna się od pionierskich roślin, tak samo firma po kryzysie tworzy nowe, elastyczne procesy, które stają się podstawą dojrzałego środowiska informacyjnego.

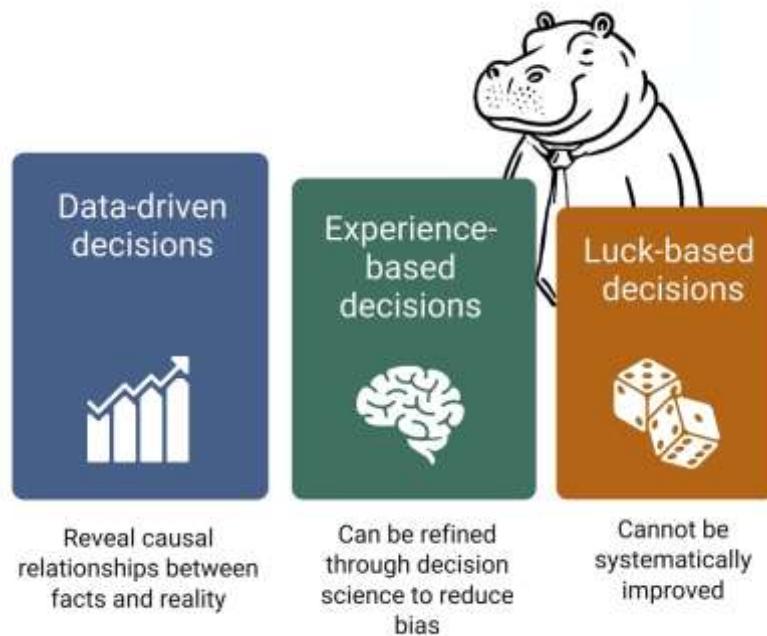
Firmy, którym uda się prawidłowo zinterpretować te "sygnały zapalne" i przekształcić ich destrukcyjną energię w konstruktywne zmiany, osiągną nowy poziom wydajności - z bardziej przejrzystymi, adaptacyjnymi procesami danych, które zwiększą naturalną zdolność organizacji do odnawiania się i rozwoju.

Rosnący wpływ sztucznej inteligencji i uczenia maszynowego na środowisko biznesowe nie budzi już wątpliwości. Nie jest to tylko tymczasowy trend, ale strategiczna konieczność. Firmy, które zignorują AI, ryzykują utratę konkurencyjności na rynku, który coraz bardziej zachęca do innowacji i elastyczności.

Przyszłość należy do tych, którzy postrzegają sztuczną inteligencję nie tylko jako narzędzie, ale jako okazję do przemyślenia każdego aspektu swojej działalności - od optymalizacji procesów po podejmowanie decyzji zarządczych.

Tworzenie cyfrowych fundamentów: 1-5 kroków do cyfrowej dojrzałości

W tym rozdziale przyjrzymy się mapie drogowej transformacji cyfrowej i zidentyfikujemy kluczowe kroki wymagane do wdrożenia podejścia opartego na danych, które może pomóc przekształcić zarówno kulturę korporacyjną, jak i ekosystem informacyjny firmy.



Rys. 10.2-2 Kontrolowana aktualizacja i wybór strategii: przypadek, doświadczenie lub dane.

Według badania McKinsey "Why Digital Strategies Fail" (2018), istnieje co najmniej pięć powodów [164], dla których firmy nie osiągają celów transformacji cyfrowej

- **Nieostre definicje:** kadra kierownicza i menedżerowie różnie rozumieją znaczenie słowa "cyfrowy", co prowadzi do nieporozumień i niespójności.
- **Niezrozumienie gospodarki cyfrowej:** wiele firm nie docenia skali zmian, jakie cyfryzacja wprowadza do modeli biznesowych i dynamiki branży (Rys. 10.1-6).
- **Ignorowanie ekosystemów:** firmy koncentrują się na indywidualnych rozwiązaniach technologicznych (silosach danych), pomijając potrzebę integracji z szerszymi ekosystemami cyfrowymi (Rys. 2.2-2, Rys. 4.1-12).
- **Niedocenianie cyfryzacji przez konkurentów:** menedżerowie nie biorą pod uwagę, że konkurenci również aktywnie wdrażają technologie cyfrowe, co może prowadzić do utraty przewagi konkurencyjnej.
- **Brak dwoistości cyfryzacji:** dyrektorzy generalni przekazują odpowiedzialność za transformację cyfrową innym członkom kadry kierowniczej, co biurokratyzuje kontrolę i spowalnia proces zmian.

Sprostanie tym wyzwaniom wymaga jasnego zrozumienia i dostosowania strategii cyfrowych na wszystkich poziomach organizacji. Przed zbudowaniem strategii cyfrowej ważne jest zrozumienie punktu wyjścia. Wiele organizacji ma tendencję do przyjmowania nowych narzędzi i platform bez pełnego obrazu

obecnego stanu.

Krok 1: Przeprowadzenie audytu obecnych systemów i danych.

Przed zmianą procesów ważne jest, aby zrozumieć, co już działa. Przeprowadzenie audytu pozwala zidentyfikować słabe punkty w zarządzaniu danymi i zrozumieć, jakie zasoby można wykorzystać. Audyt to swego rodzaju "zdjęcie rentgenowskie" procesów biznesowych. Pozwala zidentyfikować obszary ryzyka i określić, które dane są krytyczne dla projektu lub firmy, a które są drugorzędne.

Główne działania:

- Zmapuj swoje środowisko IT (w Draw.io, Lucidchart, Miro, Visio lub Canva). Wymień używane systemy (ERP, CAD, CAFM, CPM, SCM i inne) zaangażowane w Twoje procesy, które omówiliśmy w rozdziale "Technologie i systemy zarządzania w nowoczesnym budownictwie" (rys. 1.2-4)
- Ocena kwestii jakości danych dla każdego systemu pod kątem częstotliwości występowania duplikatów, możliwych brakujących wartości i niespójności formatu w każdym systemie.
- Zidentyfikuj "punkty bólu" - miejsca, w których procesy mogą się załamać lub często wymagają ręcznej interwencji - importu, eksportu i dodatkowych procesów walidacji.

Jeśli chcesz, aby zespół ufał raportom, musisz upewnić się, że dane są poprawne od samego początku.

Audit jakości danych pokaże, jakie dane są potrzebne:

- Wymaga dalszego rozwoju (należy skonfigurować automatyczne procesy czyszczenia lub dodatkowe przekształcenia).
- Są to "śmieci", które tylko zatykają systemy i można się ich pozbyć, nie używając ich już w procesach.

Możliwe jest przeprowadzenie takiego audytu we własnym zakresie. Czasami jednak warto zaangażować zewnętrznego konsultanta - zwłaszcza z innych branż: świeże spojrzenie i niezależność od "specyfiki" budowlanej pomoże trzeźwo ocenić status quo i uniknąć typowych pułapek związanych z uprzedzeniami wobec niektórych rozwiązań i technologii.

Krok 2: Określenie kluczowych standardów harmonizacji danych.

Po przeprowadzeniu audytu konieczne jest stworzenie wspólnych zasad pracy z danymi. Jak omówiliśmy w rozdziale "Standardy: od losowych plików do inteligentnego modelu danych", pomoże to wyeliminować silosowe przepływy danych.

Bez jednego standardu każdy zespół będzie nadal pracował "po swojemu", a ty będziesz utrzymywał "zoo" integracji, w którym dane są tracone przy każdej konwersji.

Główne działania:

- Wybierz standardy danych do wymiany informacji między systemami:

- W przypadku danych tabelarycznych mogą to być formaty strukturalne, takie jak CSV, XLSX lub bardziej wydajne formaty, takie jak Parquet
- Do wymiany danych i dokumentów o luźnej strukturze: JSON lub XML.

■ Mistrzowska praca z modelami danych:

- Rozpocznij od parametryzacji zadań na poziomie koncepcyjnego modelu danych - zgodnie z opisem w rozdziale "Modelowanie danych: model koncepcyjny, logiczny i fizyczny" (rys. 4.3-2).
- W miarę zagłębiania się w logikę procesu biznesowego należy przejść do formalizacji wymagań przy użyciu parametrów w modelach logicznych i fizycznych (rys. 4.3-6).
- Identyfikacja kluczowych podmiotów, ich atrybutów i relacji w ramach procesów oraz wizualizacja tych relacji - zarówno między podmiotami, jak i między parametrami (Rys. 4.3-7).

■ Używaj wyrażeń regularnych (RegEx) do walidacji i standaryzacji danych (Rys. 4.4-7), jak omówiliśmy w rozdziale "Wymagania strukturalne i wyrażenia regularne RegEx". RegEx nie jest skomplikowanym, ale niezwykle ważnym tematem w pracy nad tworzeniem wymagań na poziomie fizycznych modeli danych.

Bez standardów na poziomie danych i wizualizacji procesów niemożliwe jest zapewnienie spójnego i skalowalnego środowiska cyfrowego. Pamiętaj, że "złe dane są drogie". A koszt błędu wzrasta, gdy projekt lub organizacja staje się bardziej złożona. Ujednolicenie formatów, zdefiniowanie nazewnictwa, struktury i zasad walidacji to inwestycja w stabilność i skalowalność przyszłych rozwiązań.

Krok 3. Wdrożenie DataOps i automatyzacja procesów.

Bez dobrze zdefiniowanej architektury firmy nieuchronnie staną w obliczu rozproszonych danych zawartych w silosowych systemach informatycznych. Dane będą niezintegrowane, powielone w wielu lokalizacjach i kosztowne w utrzymaniu.

Wyobraźmy sobie, że dane to woda, a architektura danych to złożony system rurociągów, które transportują wodę ze źródła jej przechowywania do punktu użycia. To właśnie architektura danych określa, w jaki sposób informacje są gromadzone, przechowywane, przekształcane, analizowane i dostarczane do użytkowników końcowych lub aplikacji.

DataOps (Data Operations) to metodologia, która integruje gromadzenie, czyszczenie, walidację i wykorzystanie danych w jeden zautomatyzowany przepływ procesów, co szczegółowo omówiliśmy w części 8 książki.

Główne działania:

- Tworzenie i dostosowywanie przenośników ETL w celu automatyzacji procesów:
- Extract: organizuje automatyczne zbieranie danych z dokumentów PDF (rys. 4.1-2, rys. 4.1-5, rys. 4.1-7), arkuszy kalkulacyjnych Excel, modeli CAD (rys. 7.2-4), systemów ERP i innych źródeł, z którymi pracujesz.

- Przekształć: skonfiguruj automatyczne procesy przekształcania danych do jednego ustrukturyzowanego formatu i zautomatyzuj obliczenia, które będą miały miejsce poza zamkniętymi aplikacjami (rys. 7.2-8).
- Załaduj: spróbuj utworzyć automatyczne przesyłanie danych do tabel podsumowujących, dokumentów lub scentralizowanych repozytoriów (rys. 7.2-9, rys. 7.2-13, rys. 7.2-16).
- Zautomatyzuj obliczenia i procesy QTO (Quantity Take-Off), jak omówiliśmy w rozdziale "QTO Quantity Take-Off: Grupowanie danych projektu według atrybutów":
 - Skonfiguruj automatyczne wyodrębnianie objętości z modeli CAD za pomocą interfejsów API, wtyczek lub narzędzi inżynierii odwrotnej (rys. 5.2-5).
 - Utwórz reguły grupowania elementów dla różnych klas według atrybutów w formie tabel (Rys. 5.2-12).
 - Spróbuj zautomatyzować często powtarzane obliczenia objętości i kosztów poza zamkniętymi systemami modułowymi (rys. 5.2-15).
- Zaczniź używać Python i Pandas do przetwarzania danych, jak omówiliśmy w rozdziale "Python Pandas: niezastąpione narzędzie do pracy z danymi":
 - Zastosuj DataFrame do pracy z plikami XLSX i zautomatyzuj przetwarzanie danych tabelarycznych (rys. 3.4-6).
 - Automatyzacja agregacji i przekształcania informacji za pomocą różnych bibliotek Pythona
 - Użyj LLM, aby uprościć pisanie gotowych bloków kodu i całych potoków (Rys. 7.2-18).
 - Spróbuj zbudować Pipeline w Pythonie, który znajduje błędy lub dostrzega anomalie i wysyła powiadomienie do osoby odpowiedzialnej (np. kierownika projektu) (Rys. 7.4-2).

Automatyzacja oparta na zasadach DataOps pozwala przejść od ręcznej i fragmentarycznej obsługi danych do zrównoważonych i powtarzalnych procesów. Nie tylko zmniejsza to obciążenie pracowników, którzy codziennie mają do czynienia z tymi samymi przekształceniami, ale także znacznie zwiększa niezawodność, skalowalność i przejrzystość całego systemu informatycznego.

Krok 4: Stwórz otwarty ekosystem zarządzania danymi.

Pomimo rozwoju zamkniętych systemów modułowych i ich integracji z nowymi narzędziami, firmy borykają się z poważnym problemem - rosnąca złożoność takich systemów przewyższa ich użyteczność. Początkowa idea stworzenia jednej autorskiej platformy obejmującej wszystkie procesy biznesowe doprowadziła do nadmiernej centralizacji, w której wszelkie zmiany wymagają znacznych zasobów i czasu na dostosowanie.

Jak omówiliśmy w rozdziale "Korporacyjna grzybnia: jak dane łączą procesy biznesowe", skuteczne zarządzanie danymi wymaga otwartego i ujednoliconego ekosystemu, który łączy wszystkie źródła informacji.

Kluczowe elementy ekosystemu:

- Wybierz odpowiedni magazyn danych:

- W przypadku tabel i obliczeń należy korzystać z baz danych - na przykład PostgreSQL lub MySQL (rys. 3.1-7).
- W przypadku dokumentów i raportów odpowiednie może być przechowywanie w chmurze (Dysk Google, OneDrive) lub systemy obsługujące format JSON.
- Sprawdź możliwości hurtowni danych, jezior danych i innych narzędzi do scentralizowanego przechowywania i analizy dużych ilości informacji (Rys. 8.1-8).
- Wdrażanie rozwiązań umożliwiających dostęp do zastrzeżonych danych:**
 - W przypadku korzystania z zastrzeżonych systemów należy skonfigurować dostęp do nich za pośrednictwem interfejsu API lub zestawu SDK w celu uzyskania danych do przetwarzania zewnętrznego (rys. 4.1-2).
 - Zapoznanie się z możliwościami narzędzi inżynierii odwrotnej dla formatów CAD (Rys. 4.1-13).
 - Skonfiguruj linie ETL, które okresowo zbierają dane z aplikacji lub serwerów, konwertują je na otwarte formaty strukturalne i zapisują w repozytoriach (Rys. 7.2-3).
 - Omówienie w zespole, w jaki sposób zapewnić dostęp do danych bez konieczności korzystania z zastrzeżonego oprogramowania.
 - Pamiętaj: dane są ważniejsze niż interfejsy. To struktura i dostępność informacji, a nie konkretne narzędzia interfejsu użytkownika, zapewniają długoterminową wartość
- Pomyśl o stworzeniu centrum doskonałości (CoE) dla danych, jak omówiliśmy w rozdziale "Centrum doskonałości (CoE) dla modelowania danych", lub o tym, jak możesz zapewnić specjalistyczną wiedzę na temat danych w inny sposób (Rys. 4.3-9).**

Ekosystem zarządzania danymi tworzy jednolitą przestrzeń informacyjną, w której wszyscy uczestnicy projektu pracują ze spójnymi, aktualnymi i zweryfikowanymi informacjami. Jest to podstawa skalowalnych, elastycznych i niezawodnych procesów cyfrowych

Uwolnienie potencjału danych: 5-10 kroków do cyfrowej dojrzałości

Oprócz integracji technicznej, ważnym czynnikiem udanego wdrożenia rozwiązań cyfrowych jest ich przyjęcie przez użytkowników końcowych. Zaangażowanie klientów lub użytkowników w pomiar wydajności jest zarówno wyzwaniem związanym z poprawą doświadczenia użytkownika, jak i zarządzaniem zmianami w firmie. Jeśli rozwiązanie nie pasuje do znanego przepływu pracy lub nie rozwiązuje rzeczywistych problemów użytkowników lub klientów, nie będzie używane i żadna ilość dodatkowych środków i zachęt tego nie naprawi.

Transformacja jest procesem iteracyjnym opartym na analizie danych dotyczących interakcji użytkownika z nowymi procesami, z częstymi cyklami testowania, ciągłymi informacjami zwrotnymi i udoskonaleniami.

Krok 5: Budowanie kultury danych, szkolenie pracowników i zbieranie informacji zwrotnych

Nawet najbardziej zaawansowany system nie zadziała bez zaangażowania pracowników. Musisz stworzyć środowisko, w którym dane są wykorzystywane na co dzień, a zespół rozumie ich wartość.

W opublikowanym w 2024 r. raporcie brytyjskiego rządu "Data Analytics and AI in Government Project Delivery" zauważono [83], że szkolenie specjalistów posiadających niezbędne kompetencje w zakresie przetwarzania i interpretacji danych ma kluczowe znaczenie dla pomyślnego wdrożenia analizy danych i sztucznej inteligencji.

Brak specjalistycznej wiedzy w zakresie analizy danych jest jednym z kluczowych czynników ograniczających transformację cyfrową. Liderzy są przyzwyczajeni do ustalonych rutyn: cykli kwartalnych, priorytetowych inicjatyw i tradycyjnych sposobów realizacji projektów. Zmiana wymaga specjalnego lidera - wystarczająco wysokiego rangą, aby mieć wpływ, ale nie tak wysokiego, aby miał czas i motywację do prowadzenia długoterminowego projektu transformacji.

Główne działania:

- Uznanie potrzeby przejścia od subiektywnych decyzji opartych na opinii wysoko opłacanego pracownika (HiPPO) do kultury podejmowania decyzji opartej na faktach i danych, jak omówiono w rozdziale "HiPPO lub niebezpieczeństwo opinii w podejmowaniu decyzji" (Rys. 2.1-9).
- Organizowanie systematycznych szkoleń:
 - Organizuj sesje szkoleniowe na temat korzystania z ustrukturyzowanych danych i zapraszaj ekspertów z innych branż, którzy nie są uprzedzeni do produktów i koncepcji popularnych obecnie w branży budowlanej.
 - Omawianie podejść do analizy danych i narzędzi z kolegami oraz samodzielna nauka praktycznej pracy z narzędziami takimi jak Python, pandas i LLM (rys. 4.1-3, rys. 4.1-6).
 - Stworzenie biblioteki samouczków (najlepiej z krótkimi filmami) na temat strukturyzacji danych (rys. 3.2-15) i tworzenia modeli danych (rys. 4.3-6, rys. 4.3-7).
- Wykorzystanie nowoczesnych technologii edukacyjnych:
 - Wykorzystanie modeli językowych (LLM) do obsługi kodu i manipulacji danymi, w tym generowania kodu, refaktoryzacji i analizy, a także przetwarzania i interpretacji informacji tabelarycznych (Rys. 3.4-1).
 - Zbadanie, w jaki sposób kod wygenerowany przez LLM można dostosować i zintegrować z kompletnym rozwiązaniem Pipeline podczas pracy w środowisku programistycznym offline (IDE) (rys. 4.4-14, rys. 5.2-13).

Gdy menedżer nadal podejmuje decyzje w "staroświecki sposób", żadna ilość szkoleń nie przekona ludzi do poważnego traktowania analityki.

Budowanie kultury danych jest niemożliwe bez ciągłej informacji zwrotnej. Informacje zwrotne pozwalają zidentyfikować luki w procesach, narzędziach i strategiach, których nie można odkryć za pomocą wewnętrznych raportów lub formalnych wskaźników KPI. Komplementarne komentarze od użytkowników Twoich rozwiązań nie zapewnią praktycznej wartości. To krytyczne informacje zwrotne są cenne, zwłaszcza jeśli opierają się na konkretnych obserwacjach i faktach. Uzyskanie takich informacji wymaga jednak wysiłku: musisz zbudować procesy, w których uczestnicy - zarówno wewnętrzni, jak i zewnętrzni - mogą dzielić się komentarzami (może mieć sens robienie tego anonimowo) bez zniekształceń i bez obawy, że ich opinie mogą wpływać na ich własną pracę. Ważne jest, aby robili to bez zniekształceń i bez obaw o negatywne konsekwencje dla siebie.

Każde uczenie się jest ostatecznie samouczeniem się [165].

- Milton Friedman, amerykański ekonomista i statystyk

Wdrożeniu narzędzi analitycznych powinna towarzyszyć regularna weryfikacja ich skuteczności w praktyce (ROI, KPI), co można osiągnąć jedynie poprzez ustrukturyzowaną informację zwrotną od pracowników, klientów i partnerów. Pozwala to firmom nie tylko unikać powtarzania błędów, ale także szybciej dostosowywać się do zmian zachodzących w otoczeniu. Posiadanie mechanizmu gromadzenia i analizowania informacji zwrotnych jest jedną z oznak dojrzałości organizacji przechodzącej od sporadycznych inicjatyw cyfrowych do zrównoważonego modelu ciągłego doskonalenia (Rys. 2.2-5).

Krok 6: Od projektów pilotażowych do zwiększenia skali

Wybieraj bitwy na tyle duże, by miały znaczenie i na tyle małe, by je wygrać.

- Jonathan Kozol

Rozpoczęcie transformacji cyfrowej "od razu i wszędzie" jest niezwykle ryzykowne. Bardziej skutecznym podejściem jest rozpoczęcie od projektów pilotażowych i stopniowe zwiększenie skali udanych praktyk.

Główne działania:

■ Wybierz odpowiedni projekt pilotażowy:

- Zdefiniuj konkretny cel biznesowy lub proces z mierzalnymi wynikami (KPI, ROI) (rys. 7.1-5).
- Wybierz proces automatyzacji ETL, taki jak automatyczna walidacja danych lub obliczanie obciążenia (QTO) przy użyciu Python i Pandas (Rys. 5.2-10)
- Ustalenie jasnych wskaźników sukcesu (np. skrócenie czasu tworzenia specyfikacji kontroli lub raportów walidacji danych z tygodnia do jednego dnia).

■ Stosuj podejście iteracyjne:

- Zaczniź od prostych procesów konwersji danych i utwórz konwersje strumieniowe danych wieloformatowych do formatów potrzebnych w procesach (rys. 4.1-2, rys. 4.1-5).
- Stopniowo zwiększ złożoność zadań i rozszerzaj automatyzację procesów, tworząc kompletny Pipeline w IDE w oparciu o udokumentowane bloki kodu (rys. 4.1-7, rys. 7.2-18).
- Dokumentuj i nagrywaj (najlepiej za pomocą krótkich filmów) udane rozwiązania i udostępniaj je współpracownikom lub społecznościom zawodowym.

- Opracuj szablony i towarzyszącą im dokumentację w celu powielenia takich rozwiązań, aby mogły być skutecznie wykorzystywane przez współpracowników (lub członków społeczności zawodowej, w tym użytkowników mediów społecznościowych).

"Roll-up" krok po kroku pozwala utrzymać wysoką jakość zmian i nie popaść w chaos równoległych wdrożeń. Strategia "od małego do dużego" minimalizuje ryzyko i pozwala uczyć się na małych błędach, nie pozwalając im przerodzić się w krytyczne problemy.

Przejście od podejścia projektowego, w którym pracownicy są tylko częściowo zaangażowani, do tworzenia stałych zespołów (np. centrów wiedzy - CoE) pomaga zapewnić zrównoważony rozwój produktu nawet po wydaniu jego pierwszej wersji. Zespoły takie nie tylko wspierają istniejące rozwiązania, ale także stale je ulepszają.

Zmniejsza to zależność od długich zatwierdzeń: członkowie zespołu są uprawnieni do podejmowania decyzji w zakresie swoich obowiązków. W rezultacie menedżerowie są wolni od potrzeby mikrozarządzania, a zespoły mogą skupić się na tworzeniu prawdziwej wartości.

Opracowywanie nowych rozwiązań to nie sprint, ale maraton. Ci, którzy odnoszą w nim sukces, to ci, którzy początkowo koncentrują się na długoterminowej, konsekwentnej pracy.

Ważne jest, aby zdawać sobie sprawę, że technologia wymaga ciągłego rozwoju. Inwestowanie w długoterminowy rozwój rozwiązań technologicznych jest podstawą udanych operacji.

Krok 7: Korzystanie z otwartych formatów danych i rozwiązań

Jak omówiliśmy w rozdziałach dotyczących platform modułowych (ERP, PMIS, CAFM, CDE itp.), ważne jest, aby skupić się na otwartych i uniwersalnych formatach danych, które zapewniają niezależność od rozwiązań dostawców i zwiększą dostępność informacji dla wszystkich zainteresowanych stron.

Główne działania:

- Przejście od formatów zamkniętych do formatów otwartych:
 - Używaj otwartych formatów zamiast formatów zastrzeżonych lub znajdź sposób na skonfigurowanie automatycznego przesyłania lub konwersji formatów zamkniętych na formaty otwarte (Rys. 3.2-15).
 - Wdrożenie narzędzi do pracy z Parquet, CSV, JSON, XLSX, które są standardami wymiany pomiędzy większością nowoczesnych systemów (rys. 8.1-2).
 - Jeśli praca z geometrią 3D odgrywa ważną rolę w procesach, należy rozważyć użycie otwartych formatów, takich jak USD, glTF, DAE lub OBJ (Rys. 3.1-14).
- Korzystanie z wektorowych baz danych w celu wydajnej analizy i wyszukiwania informacji:
 - Użyj Bounding Box i innych metod, aby uprościć geometrię 3D (Rys. 8.2-1).
 - Zastanów się, gdzie możesz wdrożyć wektoryzację danych - konwertowanie tekstów, obiektów lub dokumentów na reprezentacje numeryczne (Rys. 8.2-2).
- Zastosowanie narzędzi do analizy dużych zbiorów danych:
 - Organizowanie przechowywania zgromadzonych danych historycznych (np. PDF, XLSX,

CAD) w formatach odpowiednich do analizy (Apache Parquet, CSV, ORC) (rys. 8.1-2).

- Rozpoczęcie stosowania podstawowych metod statystycznych i pracy z reprezentatywnymi próbками - lub przynajmniej zapoznanie się z podstawowymi zasadami statystyki (Rys. 9.2-5).
- Wdrożenie i poznanie narzędzi do wizualizacji danych i relacji między danymi w celu wizualizacji wyników analizy. Bez dobrej wizualizacji nie można w pełni zrozumieć ani samych danych, ani opartych na nich procesów (Rys. 7.1-4).

Przejście na otwarte formaty danych oraz wprowadzenie narzędzi do analizy, przechowywania i wizualizacji informacji stanowi podstawę zrównoważonego i niezależnego zarządzania cyfrowego. Nie tylko zmniejsza to zależność od dostawców, ale także zapewnia równy dostęp do danych wszystkim zainteresowanym stronom.

Krok 8. Rozpoczęcie wdrażania uczenia maszynowego do przewidywania

Wiele firm zgromadziło ogromne ilości danych - swego rodzaju "gejzery informacji", które wciąż pozostają niewykorzystane. Dane te zostały zebrane w setkach lub tysiącach projektów, ale często zostały wykorzystane tylko raz lub wcale w dalszych procesach. Dokumenty i modele przechowywane w zamkniętych formatach i systemach są często postrzegane jako przestarzały i bezużyteczny balast. W rzeczywistości są one jednak cennym zasobem - podstawą do analizy popełnionych błędów, automatyzacji rutynowych operacji i opracowywania innowacyjnych rozwiązań do automatycznej klasyfikacji i rozpoznawania cech w przyszłych projektach.

Kluczowym wyzwaniem jest nauczenie się, jak wyodrębnić te dane i przekształcić je w przydatne informacje. Jak omówiono w rozdziale poświęconym uczeniu maszynowemu i prognozowaniu, techniki uczenia maszynowego mogą znacznie poprawić dokładność szacunków i prognoz w różnych procesach związanych z budownictwem. Pełne wykorzystanie zgromadzonych danych otwiera drogę do poprawy wydajności, zmniejszenia ryzyka i budowania zrównoważonych procesów cyfrowych.

Główne działania:

- Zacznić od prostych algorytmów:
 - Spróbuj zastosować regresję liniową - korzystając z podpowiedzi LLM - do przewidywania powtarzalnej wydajności w zestawach danych, w których zależności od dużej liczby czynników są nieobecne lub minimalne (Rys. 9.3-4).
 - Zastanów się, na których etapach procesów teoretycznie można zastosować algorytm k-najbliższych sąsiadów (k-NN) - na przykład do zadań klasyfikacji, oceny podobieństwa obiektów lub przewidywania na podstawie historycznych analogów (rys. 9.3-5).
- Zbieranie i strukturyzowanie danych w celu trenowania modeli:
 - Gromadzenie historycznych danych projektu w jednym miejscu i w jednym formacie (Rys. 9.1-10).
 - Praca nad jakością i reprezentatywnością próbek szkoleniowych poprzez zautomatyzowany ETL (Rys. 9.2-8).
 - Naucz się rozdzielać dane na zestawy treningowe i testowe, tak jak zrobiliśmy to w przykładzie zestawu danych Titanic (Rys. 9.2-6, Rys. 9.2-7).

- Rozważ rozszerzenie zastosowania technik uczenia maszynowego na szeroki zakres zadań, od przewidywania harmonogramów projektów po optymalizację logistyki, zarządzanie zasobami i wczesną identyfikacją potencjalnych problemów.

Uczenie maszynowe jest narzędziem do przekształcania zarchiwizowanych danych w cenny zasób do przewidywania, optymalizacji i świadomego podejmowania decyzji. Zacznię od małych zbiorów danych (Rys. 9.2-5) i prostych modeli, stopniowo zwiększając ich złożoność.

Krok 9. Integracja IoT i nowoczesnych technologii gromadzenia danych

Świat budownictwa szybko staje się cyfrowy: każde zdjęcie z budowy, każdy post w Teams jest już częścią większego procesu parametryzacji i tokenizacji rzeczywistości. Tak jak kiedyś GPS zmienił logistykę, tak IoT, RFID i technologie automatycznego gromadzenia danych zmieniają branżę budowlaną. Jak omówiono w rozdziale "Internet rzeczy IoT i inteligentne kontrakty", cyfrowy plac budowy z czujnikami i zautomatyzowanym monitorowaniem to przyszłość branży.

Główne działania:

- Wdrożenie urządzeń IoT, tagów RFID i wyszczególnienie związanych z nimi procesów:
 - Oceń, które obszary lub etapy projektu mogą przynieść największy zwrot z inwestycji (ROI) - na przykład w przypadku monitorowania temperatury, vibracji, wilgotności lub ruchu.
 - Rozważ wykorzystanie RFID do śledzenia materiałów, narzędzi i sprzętu w całym łańcuchu dostaw.
 - Należy rozważyć, w jaki sposób zebrane dane można zintegrować w jednym systemie informacyjnym, takim jak Apache NiFi, w celu zautomatyzowanego przetwarzania i analizy w czasie rzeczywistym (Rys. 7.4-5).
- Ustanowienie systemu monitorowania w czasie rzeczywistym:
 - Opracowywanie pulpitów nawigacyjnych w celu śledzenia kluczowych wskaźników procesu lub projektu przy użyciu narzędzi do wizualizacji, takich jak Streamlit, Flask lub Power BI)
 - Skonfiguruj automatyczne powiadomienia sygnalizujące krytyczne odchylenia od planu lub norm (rys. 7.4-2).
 - Ocena potencjału konserwacji predykcyjnej sprzętu na podstawie zebranych danych i zidentyfikowanych wzorców (Rys. 9.3-6).
- Łączenie danych z różnych źródeł:
 - Zacznię od wizualizacji modelu danych na poziomie fizycznym - odzwierciedl strukturę przepływów informacji i kluczowych parametrów pochodzących z systemów CAD, urządzeń IoT i platform ERP (rys. 4.3-1).
 - Zacznię od stworzenia wstępnego zarysu ujednoliconej platformy do analizy danych i wspomagania decyzji zarządczych. Należy uchwycić kluczowe funkcje, źródła danych, użytkowników i przewidywane scenariusze zastosowań (Rys. 4.3-7).

Im szybciej zacznesz łączyć rzeczywiste procesy ze światem cyfrowym, tym szybciej będziesz mógł

zarządzać nimi za pomocą danych - wydajnie, przejrzystie i w czasie rzeczywistym.

Krok 10. Przygotuj się na przyszłe zmiany w branży

Firmy budowlane znajdują się pod ciągłą presją otoczenia zewnętrznego: kryzysów gospodarczych, skoków technologicznych, zmian regulacyjnych. Podobnie jak las, który musi wytrzymać deszcz, śnieg, suszę i palące słońce, firmy żyją w warunkach ciągłej adaptacji. I tak jak drzewa stają się odporne na mróz i suszę dzięki głębokim systemom korzeniowym, tylko te organizacje, które mają solidne podstawy zautomatyzowanych procesów, zdolność przewidywania zmian i elastycznego dostosowywania strategii, pozostają rentowne i konkurencyjne.

Jak wspomniano w rozdziale "Strategie przetrwania: budowanie przewagi konkurencyjnej", branża budowlana wchodzi w fazę radykalnej transformacji. Interakcja między klientem a wykonawcą zmierza w kierunku modelu perswazji, w którym przejrzystość, przewidywalność i narzędzia cyfrowe zastępują tradycyjne podejście. W tej nowej rzeczywistości wygrywają nie najwiękscy, ale najbardziej elastyczni i dojrzałi technologicznie.

Główne działania:

- Analiza słabych punktów w biznesie w kontekście otwartych danych:
 - Oceń, w jaki sposób demokratyzacja dostępu do danych w ramach Uberyzacji może mieć druzgocący wpływ na Twoją przewagę konkurencyjną i działalność (Rys. 10.1-5).
 - Pomyśl o strategii przejścia od nieprzejrzystych i silosowych procesów do modeli biznesowych opartych na otwartych rozwiązaniach, interoperacyjności systemów i przejrzystości danych (rys. 2.2-5).
- Opracowanie długoterminowej strategii cyfrowej:
 - Określ, czy aspirujesz do bycia liderem innowacji, czy wolisz scenariusz "nadrabiania zaległości", w którym będziesz oszczędzać swoje zasoby.
 - Opis etapów: krótkoterminowy (automatyzacja procesów, strukturyzacja danych), średnioterminowy (wdrożenie LLM i ETL), długoterminowy (ekosystemy cyfrowe, scentralizowane repozytoria).
- Pomyśl o rozszerzeniu swojego portfolio usług:
 - Rozważ oferowanie nowych usług (skoncentrowanych na efektywności energetycznej, ESG, usługach związanych z danymi). Więcej o nowych modelach biznesowych powiemy w następnym rozdziale
 - Staraj się pozycjonować jako niezawodny partner technologiczny, który wspiera cały cykl życia obiektu, od projektu po eksploatację. Zaufanie do Ciebie powinno opierać się na systematycznym podejściu, przejrzystych procesach i zdolności do dostarczania zrównoważonych rozwiązań technologicznych

W transformacyjnym środowisku wygrywają nie ci, którzy po prostu reagują na zmiany, ale ci, którzy są proaktywni. Elastyczność, otwartość i dojrzałość cyfrowa to fundamenty zrównoważonego rozwoju w budowaniu jutra.

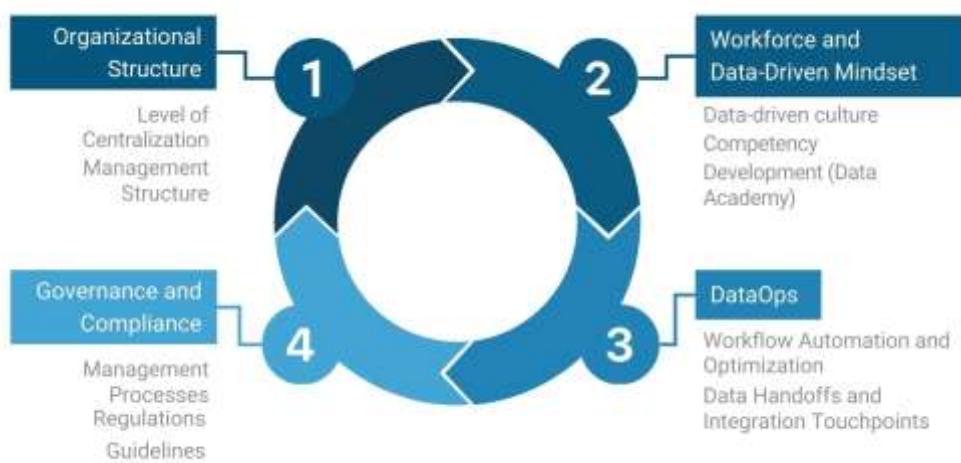
Plan transformacji: od chaosu do firmy opartej na danych

Poniższy plan może służyć jako wstępny punkt odniesienia - punkt wyjścia do kształtowania własnej strategii transformacji cyfrowej opartej na danych:

- **Audyt i standardy:** analiza obecnego stanu, ujednolicenie danych
- **Strukturyzacja i klasyfikacja danych:** automatyzacja przekształcania nieustrukturyzowanych i luźno ustrukturyzowanych danych.
- **Automatyzacja grupowania, obliczeń i kalkulacji:** wykorzystanie narzędzi i bibliotek open source do automatyzacji
- **Ekosystem i COE:** zbudowanie wewnętrznego zespołu w celu stworzenia jednolitego ekosystemu danych w firmie.
- **Kultura i nauka:** odejście od rozwiązań HiPPO na rzecz rozwiązań opartych na danych
- **Pilotaje, informacje zwrotne i skalowanie:** działaj iteracyjnie: testuj nowe metody na ograniczoną skalę, zbieraj ważne informacje zwrotne i stopniowo zwiększaj skalę rozwiązań.
- **Otwarte formaty:** korzystanie z uniwersalnych i otwartych formatów w celu uniezależnienia się od producentów oprogramowania.
- **Uczenie maszynowe:** osadzanie algorytmów uczenia maszynowego w procesach przewidywania i optymalizacji.
- **IoT i cyfrowy plac budowy:** integracja nowoczesnych technologii gromadzenia danych z procesami
- **Strategiczna adaptacja:** przygotowanie na przyszłe zmiany w branży

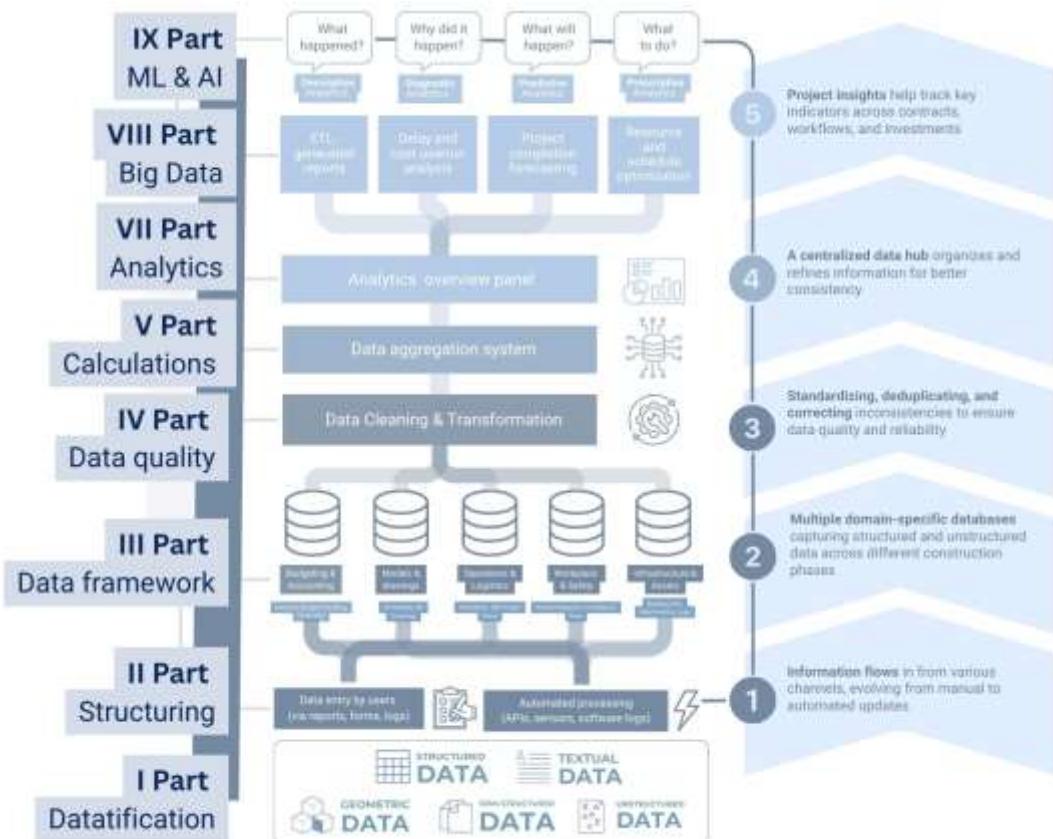
Co najważniejsze, pamiętaj, że "same dane nie zmieniają firmy: to ludzie, którzy wiedzą, jak pracować z tymi danymi, zmieniają ją". Podkreśl kulturę, przejrzyste procesy i zaangażowanie w ciągłe doskonalenie

Podejście systemowe umożliwia przejście od silosowych inicjatyw cyfrowych do pełnoprawnego modelu zarządzania opartego na danych, w którym decyzje nie są oparte na intuicji lub założeniach, ale na danych, faktach i matematycznie obliczonych prawdopodobieństwach. Cyfrowa transformacja branży budowlanej to nie tylko wdrażanie technologii, ale także kształtowanie ekosystemu biznesowego, w którym informacje o projekcie są płynnie i iteracyjnie przesyłane między różnymi systemami. W ten sposób algorytmy uczenia maszynowego zapewniają automatyczną, ciągłą analizę, prognozowanie i optymalizację procesów. W takim środowisku spekulacje i ukryte dane stają się nieistotne - pozostają tylko sprawdzone modele, przejrzyste obliczenia i przewidywalne wyniki.



Rys. 10.2-3 Kluczowe elementy skutecznego zarządzania danymi na poziomie firmy.

Każda część książki odpowiada konkretnemu etapowi przetwarzania i analizy danych w projektach budowlanych (rys. 2.2-5). Jeśli chcesz powrócić do jednego z tematów omówionych wcześniej i spojrzeć na niego z perspektywy całościowego zrozumienia przepływu wykorzystania danych, możesz odwołać się do tytułów części na rys. 10.2-4.



Rys. 10.2-4 Części książki w kontekście potoku przetwarzania danych (rys. 2.2-5): od cyfryzacji informacji do analityki i sztucznej inteligencji.

Niezależnie od wielkości organizacji, poziomu dojrzałości technologicznej czy budżetu, już dziś można zacząć stosować podejście oparte na danych. Nawet małe kroki we właściwym kierunku przyniosą z czasem rezultaty.

Transformacja oparta na danych nie jest jednorazowym projektem, ale ciągłym, iteracyjnym procesem doskonalenia, który obejmuje wprowadzanie nowych narzędzi, przeprojektowywanie procesów i rozwijanie kultury podejmowania decyzji opartych na danych.

Budowanie w Przemyśle 5.0: Jak zarabiać pieniądze, gdy nie można się już ukryć?

Przez długi czas firmy budowlane zarabiały na nieprzejrzystości procesów. Głównym modelem biznesowym była spekulacja - zawyżanie kosztów materiałów, zakresu prac i narzutów procentowych w zamkniętych systemach ERP -, PMIS - niedostępnych dla audytu zewnętrznego. Ograniczony dostęp klientów i ich zaufanych osób do oryginalnych danych projektowych stworzył grunt dla schematów, w których weryfikacja wiarygodności obliczeń stała się niemal niemożliwa.

Model ten jednak szybko traci na znaczeniu. Wraz z demokratyzacją dostępu do danych, pojawiением się LLM, pojawiением się otwartych danych i narzędzi do automatyzacji ETL, branża przechodzi do nowego standardu pracy.

W rezultacie nieprzejrzystość nie jest już przewagą konkurencyjną - wkrótce stanie się ciężarem, z którym trudno będzie się rozstać. Przejrzystość zmienia się z opcji w warunek konieczny do utrzymania się na rynku.

Z kim będą współpracować klienci - banki, inwestorzy, klienci fizyczni, private equity, klienci rządowi - w nowej cyfrowej rzeczywistości? Odpowiedź jest oczywista: z tymi, którzy są w stanie zapewnić nie tylko wynik, ale także uzasadnienie każdego kroku na drodze do niego. Wraz ze wzrostem ilości otwartych danych, partnerzy i klienci będą wybierać firmy, które gwarantują przejrzystość, dokładność i przewidywalność wyników.

W tym kontekście pojawiają się nowe modele biznesowe, które opierają się na zarządzaniu danymi i zaufaniu, a nie na spekulacjach:

■ **Sprzedaż procesów zamiast metrów kwadratowych:** kluczowym atutem staje się zaufanie i wydajność zamiast zdyskontowanych konkretnych umów. Główną wartością będzie przewidywalność wyniku oparta na wiarygodnych i zweryfikowanych danych. Nowoczesne firmy nie będą sprzedawać obiektu budowlanego jako takiego, ale:

- dokładne terminy i przejrzyste harmonogramy pracy;
- rozsądne szacunki, poparte obliczeniami;
- Pełna cyfrowa identyfikowalność i kontrola na wszystkich etapach projektu.

■ **Inżynieria i analityka jako usługa:** model "Data-as-a-Service " (sposób dostarczania gotowych danych użytkownikom za pośrednictwem Internetu, jako usługa), w którym każdy projekt staje się częścią cyfrowego łańcucha danych, a wartość biznesowa polega na umiejętności zarządzania tym łańcuchem. Firmy przekształcają się w inteligentne platformy oferujące rozwiązania oparte na automatyzacji i analityce:

- Zautomatyzowane i przejrzyste przygotowywanie szacunków i planów;
- ocena ryzyka i czasu w oparciu o algorytmy uczenia maszynowego;
- obliczanie wskaźników środowiskowych (ESG, CO₂, efektywność energetyczna);
- generowanie raportów z audytowanych źródeł otwartych.

■ **Produktywizacja wiedzy inżynierskiej:** rozwój firmy może być wielokrotnie wykorzystywany w

firmie i dystrybuowany jako oddzielny produkt - tworząc dodatkowe źródło dochodu dzięki usługom cyfrowym. W nowym środowisku firmy tworzą nie tylko projekty, ale także zasoby cyfrowe:

- biblioteki komponentów i szablony kosztorysów;
- moduły automatycznej weryfikacji;
- Wtyczki i skrypty open-source (sprzedaż konsultingowa) do pracy z danymi.

■ **Nowy typ firmy: Data-Driven Integrator:** gracz rynkowy, który nie jest zależny od konkretnych dostawców oprogramowania lub systemów modułowych i nie jest "zamknięty" w jednym interfejsie oprogramowania. Działa swobodnie z danymi - i na nich buduje swoją konkurencyjność. Firma budowlana przyszłości to nie tylko wykonawca, ale integrator informacji zdolny do wykonywania następujących funkcji dla klienta:

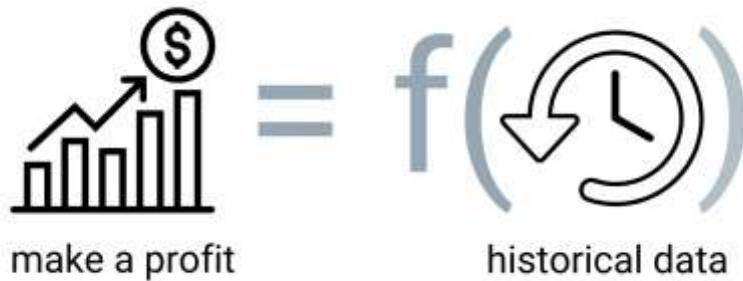
- Łączenie danych z różnych źródeł i przeprowadzanie analiz;
- Zapewnienie przejrzystości i wiarygodności procesów;
- doradztwo w zakresie optymalizacji procesów biznesowych;
- rozwijać narzędzia, które działają w ekosystemie otwartych danych, LLM, ETL i Pipelines.

Przemysł 5.0 (rys. 2.1-12) oznacza koniec "ery ręcznego uśredniania wskaźników" i wieczornych spotkań prezesów z działem szacunków i księgowości. Wszystko, co wcześniej było ukryte - obliczenia, szacunki, volumeny - staje się jawne, weryfikowalne i zrozumiałe nawet dla osób niebędących ekspertami. Ci, którzy jako pierwsi dokonają reorientacji, będą zwycięzcami. Wszyscy inni zostaną wykluczeni z nowej cyfrowej gospodarki sektora budowlanego.

WNIOSKI

Branża budowlana wkracza w erę fundamentalnych zmian. Od pierwszych zapisów na glinianych tabliczkach po ogromne ilości danych cyfrowych przepływających z serwerów projektowych i placów budowy, historia zarządzania informacjami w budownictwie zawsze odzwierciedlała dojrzałość technologii swoich czasów. Obecnie, wraz z pojawiением się automatyzacji, otwartych formatów i inteligentnych systemów analitycznych, branża stoi w obliczu nie stopniowej ewolucji, ale szybkiej cyfrowej transformacji.

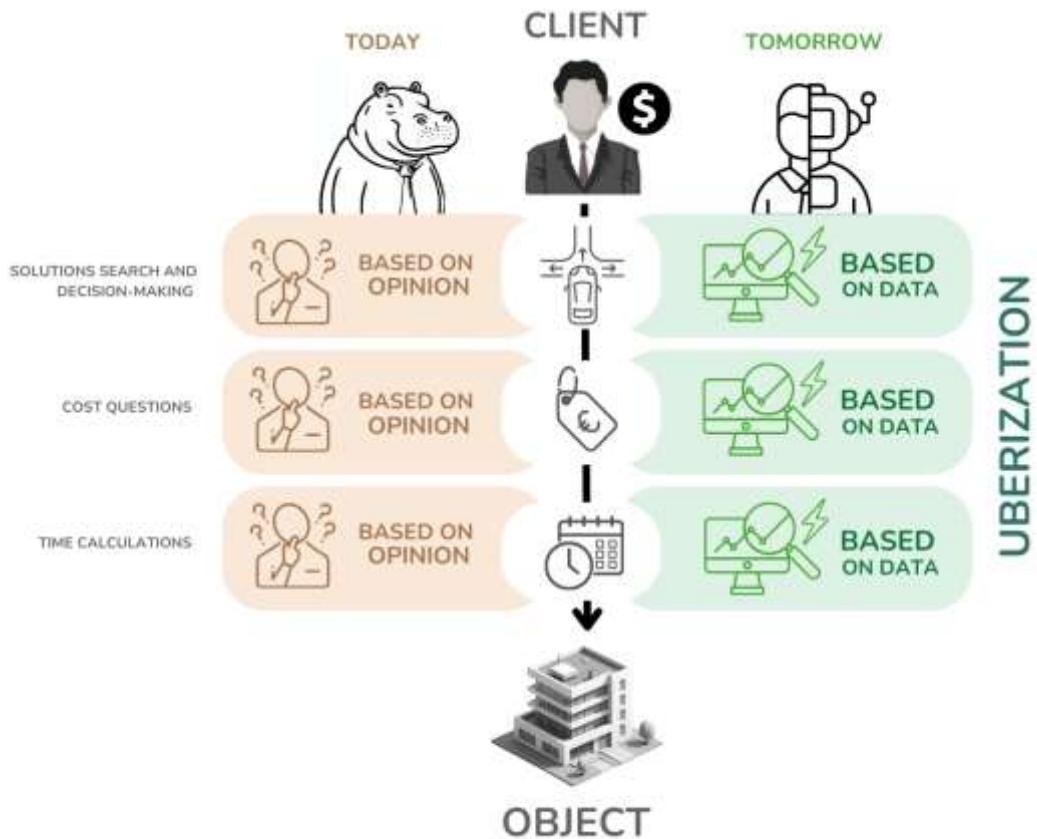
Podobnie jak w innych sektorach gospodarki, budownictwo będzie musiało przemyśleć nie tylko narzędzia, ale także zasady pracy. Firmy, które do tej pory dyktowały warunki na rynku i pełniły rolę głównego pośrednika między klientem a projektem, tracą swoją wyjątkową pozycję. Na pierwszy plan wysuwa się zaufanie i umiejętność pracy z danymi - od ich gromadzenia i strukturyzowania po analizę, prognozowanie i automatyzację decyzji.



Rys. 10.2-1 Ustrukturyzowane dane historyczne są paliwem dla wydajnej i łatwej w zarządzaniu firmy.

Niniejsza książka szczegółowo opisuje kluczowe zasady zarządzania danymi w branży budowlanej - od audytu i standaryzacji po automatyzację procesów, wykorzystanie narzędzi do wizualizacji i wdrażanie inteligentnych algorytmów. Przyjrzyliśmy się, w jaki sposób, nawet przy ograniczonych zasobach, można zbudować działającą architekturę danych i zacząć podejmować decyzje w oparciu o weryfikowalne fakty, a nie intuicję. Praca z danymi nie jest już tylko zadaniem działu IT - staje się fundamentem kultury zarządzania, która determinuje elastyczność, zdolność adaptacji i długoterminową stabilność firmy.

Zastosowanie technologii uczenia maszynowego, systemów automatycznego przetwarzania, cyfrowych bliźniaków i otwartych formatów już dziś umożliwia wyeliminowanie czynnika ludzkiego tam, gdzie wcześniej miał on kluczowe znaczenie. Budownictwo zmierza w kierunku autonomii i sterowalności, gdzie przejście od pomysłu do realizacji projektu można porównać do nawigacji w trybie autopilotu: bez zależności od subiektywnych decyzji, bez potrzeby ręcznej interwencji na każdym etapie, ale z pełną cyfrową identyfikowalnością i kontrolą (rys. 10.2-2).



Rys. 10.2-2 Przejście od podejmowania decyzji w oparciu o opinię ważnych ekspertów (HiPPO) do analizy danych będzie promowane przede wszystkim przez klienta.

Zapoznając się z metodami, zasadami i narzędziami przedstawionymi w tej książce, będziesz w stanie zacząć podejmować w swojej firmie decyzje oparte na danych, a nie na intuicji. Będziesz także w stanie uruchamiać łańcuchy modułów w LLM, kopiować gotowe potoki ETL do swojego środowiska programistycznego (IDE) i automatycznie przetwarzając dane, aby uzyskać potrzebne informacje w żądanej formie. Opierając się na rozdziałach książki poświęconych big data i uczeniu maszynowemu, będziesz w stanie wdrożyć bardziej złożone scenariusze - wydobywając nowe spostrzeżenia z danych historycznych i stosując algorytmy uczenia maszynowego do przewidywania i optymalizacji procesów.

Otwarte dane i procesy zapewnią podstawę do dokładniejszych szacunków kosztów i harmonogramów projektów, uniemożliwiając firmom budowlanym spekulowanie na podstawie nieprzejrzystych danych. Jest to zarówno wyzwanie, jak i szansa dla branży na ponowne przemyślenie swojej roli i dostosowanie się do nowego środowiska, w którym przejrzystość i wydajność staną się kluczowymi czynnikami sukcesu.

Chęć wykorzystania wiedzy i zastosowania jej w praktyce jest kluczem do sukcesu w erze cyfrowej transformacji.

Firmy, które jako pierwsze zdadzą sobie z tego sprawę, będą miały przewagę w nowej cyfrowej konkurencji. Ważne jest jednak, aby zdać sobie sprawę, że same dane niczego nie zmieniają. Wiele osób będzie musiało zmienić sposób myślenia, a to wymaga zachęty. Twój biznes musi przemyśleć sposób, w jaki udostępnia dane

Ludzie, którzy zmieniają firmę, to ludzie, którzy wiedzą, jak pracować z tymi danymi, interpretować je, wykorzystywać do optymalizacji i tworzyć na ich podstawie nową architekturę procesów.

Jeśli czytasz te słowa, jesteś gotowy na zmiany i już jesteś o krok do przodu. Dziękujemy za wybranie tej ścieżki. Witamy w erze cyfrowej transformacji!



O AUTORZE

Nazywam się Artyom Boiko. Moja podróż na placu budowy rozpoczęła się w 2007 roku - od pracy jako górnik w kopalni łupków naftowych w moim rodzinnym mieście, podczas studiów na Uniwersytecie Górniczym w Sankt Petersburgu, specjalizującym się w kopalniach i budownictwie podziemnym. Na tyłnej okładce tej książki można zobaczyć eksplozjonistę w przodku, w którym wydobywaliśmy i wysadzaliśmy setki kostek łupków naftowych. Moja kariera prowadziła mnie w wielu różnych kierunkach, od pracy jako górnik i pracownik budownictwa podziemnego po alpinistę przemysłowego, dekarza i instalatora sprzętu dźwigowego. Miałem zaszczyt uczestniczyć w projektach o różnej skali, od budowy domów prywatnych po duże obiekty przemysłowe w różnych regionach świata.



Z biegiem czasu moja praca przeniosła się z fizycznego budownictwa na zarządzanie informacjami i procesy cyfrowe. Od 2013 roku pracowałem na różnych stanowiskach w małych, średnich i dużych firmach budowlanych w kilku regionach Niemiec, od planisty do kierownika ds. zarządzania danymi. Jeśli chodzi o zarządzanie danymi, moje doświadczenie obejmuje pracę z danymi w różnych systemach ERP, CAD (BIM), MEP, FEM, CMS. Zajmowałem się optymalizacją, automatyzacją procesów, a także analizą, uczeniem maszynowym, przetwarzaniem danych w fazach planowania, obliczeń i realizacji robót budowlanych w przedsiębiorstwach budownictwa przemysłowego, mieszkaniowego, infrastrukturalnego i użyteczności publicznej.

Pracuję z otwartym oprogramowaniem i otwartymi danymi od 2003 roku. W tym czasie wdrożyłem wiele projektów internetowych - od stron internetowych i sklepów internetowych po pełnoprawne aplikacje internetowe wykorzystujące rozwiązania open source i otwarte systemy CMS. Platformy te, podobne pod wieloma względami do nowoczesnych systemów ERP, mają modułową architekturę, wysoką zdolność adaptacji i dostępność. To doświadczenie położyło podwaliny pod moje profesjonalne podejście - skupienie się na technologiach open source i kulturze wspólnego rozwoju. Szacunek dla otwartego oprogramowania i swobodnej wymiany wiedzy jest czymś, co staram się promować w branży budowlanej. Moja praca na rzecz poprawy dostępności danych w branży budowlanej przełożyła się na stworzenie kilku społeczności mediów społecznościowych w celu omówienia otwartości danych i wykorzystania Open Source w budownictwie, a także uruchomienie kilku startupów opracowujących rozwiązania zapewniające dostęp do danych z różnych zamkniętych systemów i platform.

Mój wkład w społeczność zawodową polega na uczestnictwie jako prelegent w konferencjach dotyczących interoperacyjności CAD (BIM), ERP, 4D-5D, LLM Machine Learning i sztucznej inteligencji, a także artykułów publikowanych w europejskich publikacjach branży budowlanej. Jednym z moich znaczących osiągnięć jest stworzenie "Historii BIM" [111], kompleksowej mapy ważnych rozwiązań programowych do zarządzania danymi w branży budowlanej. Moja siedmioczęściowa seria artykułów "BIM Development and Lobbying Games", przetłumaczona na kilka języków, została powszechnie uznana za próbę naświetlenia ukrytej dynamiki standardów cyfrowych.

W ten sposób przeszedłem od wydobywania skał do wydobywania i systematyzowania danych budowlanych. Zawsze jestem otwarty na profesjonalny dialog, nowe pomysły i wspólne projekty. Z wdzięcznością przyjmę każdą informację zwrotną i czekam na Twoje wiadomości lub spotkanie z Tobą wśród moich obserwatorów w mediach społecznościowych. Bardzo dziękuję za przeczytanie tej książki do końca! Będę szczęśliwy, jeśli ta książka pomoże Ci lepiej zrozumieć temat danych w branży budowlanej.

DALSZE RELACJE

Opinie czytelników odgrywają ważną rolę w dalszym rozwoju publikacji i wyborze priorytetowych tematów. Szczególnie cenne są komentarze dotyczące tego, które pomysły okazały się przydatne, a które wzbudziły wątpliwości i wymagają dalszych wyjaśnień lub przytoczenia źródeł. Książka zawiera szeroki zakres materiałów i analiz, z których część może wydawać się kontrowersyjna lub subiektywna. Jeśli w trakcie lektury natrafisz na nieścisłości, błędnie przytoczone źródła, niespójności logiczne lub literówki, będę wdzięczny za uwagi, przemyślenia lub krytykę, które możesz przesłać na adres: boikoartem@gmail.com. Lub za pośrednictwem wiadomości na LinkedIn: linkedin.com/in/boikoartem

Będę wdzięczny za wszelkie wzmianki o książce Data-Driven Construction w mediach społeczeństwowych - dzielenie się wrażeniami z lektury pomaga rozpowszechniać informacje o otwartych danych i narzędziach oraz wspiera moją pracę.

KOMENTARZ DO TŁUMACZENIA

Ta książka została przetłumaczona przy użyciu technologii sztucznej inteligencji. Znacznie przyspieszyło to proces tłumaczenia. Jednakże, jak w przypadku każdej operacji technologicznej, mogą wystąpić błędy lub nieścisłości. Jeśli zauważysz coś, co wydaje się niepoprawne lub nieprawidłowo przetłumaczone, napisz do mnie. Twoje uwagi pomogą poprawić jakość tłumaczenia.

SPOŁECZNOŚCI BUDOWLANE OPARTE NA DANYCH

To miejsce, w którym możesz swobodnie zadawać pytania i dzielić się swoimi problemami i rozwiązaniami:

DataDrivenConstruction.io: <https://datadrivenconstruction.io>

LinkedIn: <https://www.linkedin.com/company/datadrivenconstruction/>

Twitter: <https://twitter.com/datadrivenconst>

Telegram: <https://t.me/datadrivenconstruction>

YouTube: <https://www.youtube.com/@datadrivenconstruction>

INNE UMIEJĘTNOŚCI I POJĘCIA

Oprócz kluczowych zasad pracy z danymi w branży budowlanej, DataDrivenConstruction porusza szeroki zakres dodatkowych koncepcji, programów i umiejętności, które są niezbędne dla profesjonalisty opartego na danych. Niektóre z nich zostały przedstawione jedynie w ogólnym zarysie, ale mają kluczowe znaczenie dla praktyki.

Zainteresowany czytelnik może odwiedzić stronę DataDrivenConstruction.io, gdzie znajdują się linki do dodatkowych materiałów dotyczących kluczowych umiejętności. Materiały te obejmują pracę z Pythonem i Pandas, budowanie procesów ETL, przykłady przetwarzania danych w projektach budowlanych CAD, systemy big data oraz nowoczesne podejście do wizualizacji i analizy danych budowlanych.

W przygotowaniu książki "DataDrivenConstruction" i wszystkich studiów przypadku wykorzystano wiele narzędzi i oprogramowania typu open source. Autor chciałby podziękować twórcom i współautorom następujących rozwiązań:

- Python i Pandas - podstawa manipulacji i automatyzacji danych
- Scipy, NumPy, Matplotlib i Scikit-Learn - biblioteki do analizy danych i uczenia maszynowego.
- SQL i Apache Parquet - narzędzia do przechowywania i przetwarzania dużych ilości danych konstrukcyjnych
- Open Source CAD (BIM) otwarte narzędzia danych w otwartych formatach
- N8n, Apache Airflow, Apache NiFi - systemy orkiestracji i automatyzacji przepływu pracy
- DeepSeek, LlaMa, Mistral - Open Source LLM

Specjalne podziękowania dla wszystkich uczestników dyskusji na temat otwartych danych i narzędzi w społecznościach zawodowych i sieciach społecznościowych, których krytyka, komentarze i pomysły pomogły ulepszyć treść i strukturę tej książki.

Śledź rozwój projektu na stronie DataDrivenConstruction.io, gdzie publikowane są nie tylko aktualizacje książki i poprawki, ale także nowe rozdziały, samouczki i praktyczne przykłady zastosowania opisanych technik.

MAKSYMALNA WYGODA DZIĘKI WERSJI DRUKOWANEJ

Posiadasz darmową wersję cyfrową **Data-Driven Construction**. W celu wygodniejszej pracy i szybkiego dostępu do materiałów zalecamy zwrócenie uwagi na **wydanie drukowane**:



■ **Zawsze pod ręką:** książka w formacie drukowanym stanie się niezawodnym narzędziem pracy, pozwalającym szybko znaleźć i wykorzystać odpowiednie wizualizacje i diagramy w każdej sytuacji roboczej.

■ **Wysokiej jakości ilustracje:** wszystkie obrazy i grafiki w drukowanym wydaniu są prezentowane w maksymalnej jakości.

■ **Szybki dostęp do informacji:** łatwa nawigacja, możliwość tworzenia notatek, zakładek i pracy z książką w dowolnym miejscu.

Kupując pełną drukowaną wersję książki, otrzymujesz wygodne narzędzie do wygodnej i efektywnej pracy z informacją: możliwość szybkiego korzystania z materiałów wizualnych w codziennych zadaniach, szybkiego wyszukiwania potrzebnych schematów i robienia notatek. Ponadto zakup wspiera rozpowszechnianie otwartej wiedzy.

Zamów drukowaną wersję książki na stronie: datadrivenconstruction.io/books



WYJĄTKOWA OKAZJA DO STRATEGICZNE POZYCJONOWANIE

Oferujemy możliwość zamieszczania materiałów reklamowych w bezpłatnej wersji DataDrivenConstruction. Płatna wersja publikacji przyciągnęła uwagę specjalistów z ponad 50 krajów świata - od Ameryki Łacińskiej po region Azji i Pacyfiku - już w pierwszym roku po jej opublikowaniu. Aby omówić indywidualne warunki współpracy i uzyskać szczegółowe informacje na temat możliwości umieszczenia, prosimy o wypełnienie formularza zwrotnego na oficjalnym portalu datadrivenconstruction.io lub napisanie do osób kontaktowych wymienionych na końcu książki.



ROZDZIAŁY KSIĄŻKI SĄ DOSTĘPNE NA STRONIE DATADRIVENCONSTRUCTION.IO

Rozdziały Data-Driven Construction można przeczytać na stronie internetowej Data-Driven Construction, na której stopniowo publikowane są kolejne części książki, dzięki czemu można szybko znaleźć potrzebne informacje i wykorzystać je w swojej pracy. Znajdziesz tam również wiele innych publikacji na podobne tematy, a także przykłady zastosowań i rozwiązań, które pomogą Ci rozwinąć umiejętności i zastosować dane w budownictwie.



NAJNOWSZE WERSJE KSIĄŻKI DO POBRANIA Z OFICJALNEJ STRONY INTERNETOWEJ

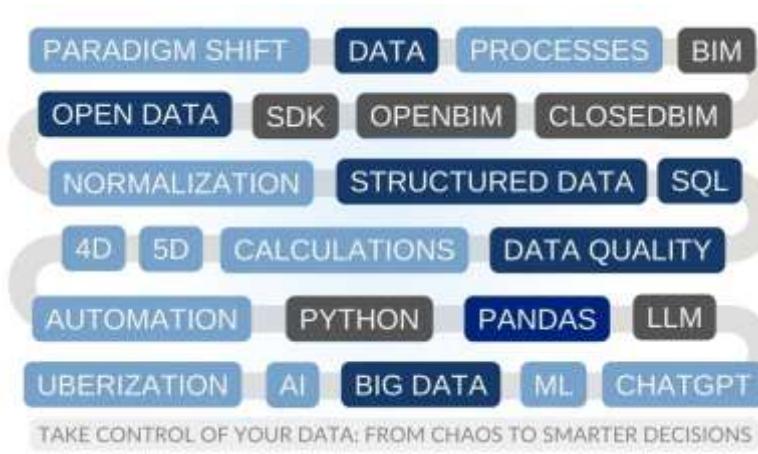
Aktualne i najnowsze wersje książki DataDrivenConstruction są dostępne do pobrania na stronie datadrivenconstruction.io. Jeśli chcesz otrzymywać aktualizacje z nowymi rozdziałami książki, praktycznymi wskazówkami lub recenzjami nowych aplikacji, zapisz się do newslettera:

- Będziesz pierwszym, który zapozna się z nowymi sekcjami książki
- Praktyczne studia przypadków i porady dotyczące analityki i automatyzacji w branży budowlanej
- Bądź na bieżąco z trendami, publikacjami i przykładami zastosowań

Przejdź do datadrivenconstruction.io, aby zasubskrybiwać!

DATA DRIVEN CONSTRUCTION: DORADZTWO, WARSZTATY I SZKOLENIA

Programy szkoleniowe i doradztwo DataDrivenConstruction pomogły dziesiątkom wiodących firm budowlanych na całym świecie zwiększyć wydajność, obniżyć koszty i poprawić jakość rozwiązań. Do klientów DataDrivenConstruction należą jedni z największych graczy na wartym miliardy euro rynku, w tym firmy budowlane, konsultingowe i informatyczne.



Dlaczego warto nas wybrać?

- **Istotność:** mówienie o głównych trendach i spostrzeżeniach dotyczących branży.
- **Praktyczne:** pomoc profesjonalistom w skutecznym rozwiązywaniu codziennych zadań poprzez PoC.
- **Spersonalizowane podejście:** dostosowane do Twojej firmy, maksymalizujące korzyści płynące ze szkoleń i doradztwa.

Główne obszary zainteresowania zespołu DataDrivenConstruction:

- **Zarządzanie jakością danych:** pomoc w parametryzacji zadań, zbieranie wymagań, walidacja i przygotowanie danych do automatycznego przetwarzania.
- **Data Mining - wyodrębnianie i strukturyzowanie danych:** procesów ETL i wyodrębnianie danych z wiadomości e-mail, plików PDF, Excel, obrazów i innych źródeł.
- **Analityka BIM i CAD :** zbieranie, strukturyzowanie i analizowanie informacji z plików RVT, IFC, DWG i innych formatów CAD (BIM).
- **Analityka i transformacja danych:** przekształcanie rozproszonych informacji w ustrukturyzowane dane, analizy, spostrzeżenia i decyzje.
- **Integracja danych i automatyzacja procesów:** od automatycznego tworzenia dokumentów po integrację z systemami wewnętrznymi i zewnętrznymi bazami danych.

Skontaktuj się z DataDrivenConstruction.io, aby dowiedzieć się, w jaki sposób automatyzacja może pomóc Twojej firmie osiągnąć wymierne wyniki biznesowe.

SŁOWNICZEK

AI (Artificial Intelligence) - sztuczna inteligencja; zdolność systemów komputerowych do wykonywania zadań, które normalnie wymagają ludzkiej inteligencji, takich jak rozpoznawanie wzorców, uczenie się i podejmowanie decyzji.

Apache Airflow to platforma orkiestracji przepływu pracy typu open source, która umożliwia programowe tworzenie, planowanie i śledzenie przepływów pracy i ETL przy użyciu DAG (skierowanych grafów acyklicznych).

Apache NiFi to narzędzie do automatyzacji przepływu danych między systemami, specjalizujące się w routingu i transformacji danych.

Apache Parquet to wydajny format plików do kolumnowego przechowywania danych, zoptymalizowany do użytku w systemach analizy dużych zbiorów danych. Zapewnia on znaczną kompresję i szybkie przetwarzanie.

API (Application Programming Interface) - sformalizowany interfejs, który pozwala jednemu programowi na interakcję z innym bez dostępu do kodu źródłowego, wymieniając dane i funkcje za pomocą standardowych żądań i odpowiedzi.

Atrybut - cecha lub właściwość obiektu, która opisuje jego cechy (np. obszar, objętość, koszt, materiał).

Bazy danych to zorganizowane struktury służące do przechowywania, zarządzania i uzyskiwania dostępu do informacji, wykorzystywane do wydajnego wyszukiwania i przetwarzania danych.

BEP (BIM Execution Plan) - plan wdrożenia modelowania informacji o budynku, który określa cele, metody i procesy wdrażania BIM w projekcie.

Big Data - tablice informacji o znacznej objętości, różnorodności i szybkości aktualizacji, wymagające specjalnych technologii do przetwarzania i analizy.

BI (Business Intelligence) - Business **Intelligence**; procesy, technologie i narzędzia do przekształcania danych w znaczące informacje do podejmowania decyzji.

BIM (Building Information Modeling) - modelowanie informacji o budynku; proces tworzenia i zarządzania cyfrowymi reprezentacjami fizycznych i funkcjonalnych cech projektów budowlanych, w tym nie tylko modeli 3D, ale także informacji o wydajności, materiałach, harmonogramie i kosztach.

BlackBox/WhiteBox - podejścia do zrozumienia systemu: w pierwszym przypadku wewnętrzna logika jest ukryta, widoczne są tylko wejścia i wyjścia; w drugim przypadku przetwarzanie jest przejrzyste i dostępne do analizy.

Bounding Box to konstrukcja geometryczna, która opisuje granice obiektu w przestrzeni trójwymiarowej za pomocą minimalnych i maksymalnych współrzędnych X, Y i Z, tworząc "pułkot" wokół obiektu.

BREP (Boundary Representation) to geometryczna reprezentacja obiektów, która definiuje je poprzez granice powierzchni.

CAD (Computer-Aided Design) to system komputerowego wspomagania projektowania używany do tworzenia, edycji i analizy dokładnych rysunków i modeli 3D w architekturze, budownictwie, inżynierii i innych branżach.

CAFM (Computer-Aided Facility Management) to oprogramowanie do zarządzania nieruchomościami i infrastrukturą, które obejmuje planowanie przestrzeni, zarządzanie aktywami, konserwację i monitorowanie kosztów.

CDE (Common Data Environment) - skoncentrowana przestrzeń cyfrowa do zarządzania, przechowywania, udostępniania i współpracy z informacjami projektowymi na wszystkich etapach cyklu życia obiektu.

Centrum Doskonałości (Centre of Excellence - CoE) to wyspecjalizowana struktura w ramach organizacji odpowiedzialna za rozwój określonego obszaru wiedzy, opracowywanie standardów i najlepszych praktyk, szkolenie personelu i wspieranie wprowadzania innowacji.

CoClass to nowoczesny system klasyfikacji elementów budowlanych trzeciej generacji.

Pojęciowy model danych to wysokopoziomowa reprezentacja podstawowych encji i ich relacji bez szczegółów atrybutów, wykorzystywana na początkowych etapach projektowania bazy danych.

CRM (Customer Relationship Management) to system zarządzania relacjami z klientami wykorzystywany do automatyzacji procesów sprzedaży i obsługi.

DAG (Directed Acyclic Graph) to skierowany graf acykliczny używany w systemach orkiestracji danych (Airflow, NiFi) do określania sekwencji zadań i zależności.

Dash jest frameworkm Pythona służącym do tworzenia interaktywnych, internetowych wizualizacji danych.

Dashboard - pulpit nawigacyjny, który wizualnie prezentuje kluczowe wskaźniki wydajności i metryki w czasie rzeczywistym.

Podejście Data-Centric to metodologia, w której dane mają pierwszeństwo przed aplikacjami lub kodem oprogramowania, dzięki czemu dane stają się centralnym zasobem organizacji.

Zarządzanie danymi - zestaw praktyk, procesów i zasad, które zapewniają odpowiednie i skuteczne wykorzystanie danych w organizacji, w tym kontrolę dostępu, jakości i bezpieczeństwa.

Data Lake to magazyn przeznaczony do przechowywania dużych ilości nieprzetworzonych danych w ich oryginalnym formacie do czasu ich wykorzystania.

Data Lakehouse to podejście architektoniczne, które łączy elastyczność i skalowalność jezior danych (Data Lake) z łatwością zarządzania i wydajnością hurtowni danych (DWH).

Data-Driven Construction to strategiczne podejście, w którym każdy etap cyklu życia obiektu - od projektowania po eksplotację - jest wspierany przez zautomatyzowane, wzajemnie połączone systemy. Takie podejście zapewnia ciągłe, oparte na faktach uczenie się, zmniejsza niepewność i umożliwia firmom osiągnięcie trwałej pozycji lidera w branży.

Data-Driven Integrator - firma specjalizująca się w łączeniu danych z różnych źródeł i analizowaniu ich w celu podejmowania decyzji zarządczych.

Podejście oparte na danych - metodologia, w której dane są postrzegane jako zasób strategiczny, a decyzje są podejmowane na podstawie obiektywnej analizy informacji, a nie subiektywnych opinii.

Minimalizm danych - podejście polegające na redukcji danych do najbardziej wartościowych i znaczących, pozwalające na uproszczone przetwarzanie i analizę informacji.

Bagno danych - rozproszona masa nieustrukturyzowanych danych, która pojawia się, gdy informacje są gromadzone i przechowywane w niekontrolowany sposób bez odpowiedniej organizacji.

DataOps to metodologia, która łączy zasady DevOps, dane i analitykę, koncentrując się na poprawie współpracy, integracji i automatyzacji przepływów danych.

Digitalizacja informacji to proces przekształcania wszystkich aspektów działalności budowlanej w formę cyfrową nadającą się do analizy, interpretacji i automatyzacji.

DataFrame - dwuwymiarowa tabelaryczna struktura danych w bibliotece Pandas, w której wiersze reprezentują poszczególne rekordy lub obiekty, a kolumny reprezentują ich cechy lub atrybuty.

Analiza opisowa - analiza danych historycznych w celu zrozumienia tego, co wydarzyło się w przeszłości.

Analityka diagnostyczna - analiza danych w celu ustalenia, dlaczego coś się wydarzyło.

Wykres Gantta to narzędzie do planowania projektu, które przedstawia zadania jako poziome paski na osi czasu, umożliwiając wizualizację sekwencji i czasu trwania pracy.

DWH (Data Warehouse) to skoncentrowany system hurtowni danych, który aggrejuje informacje z wielu źródeł, strukturyzuje je i udostępnia do analizy i raportowania.

ESG (Environmental, Social, Governance) - zestaw kryteriów oceny wpływu spółki lub projektu na środowisko, społeczeństwo i ład korporacyjny.

ELT (Extract, Load, Transform) to proces, w którym dane są najpierw pobierane ze źródeł i ładowane do repozytorium, a następnie przekształcane do celów analitycznych.

ETL (Extract, Transform, Load) to proces wyodrębniania danych z różnych źródeł, przekształcania ich do pożądanego formatu i ładowania do docelowej pamięci masowej w celu analizy.

Diagram ER (Entity-Relationship) - diagram wizualny przedstawiający encje, ich atrybuty i relacje między nimi, używany w modelowaniu danych.

ERP (Enterprise Resource Planning) to kompleksowy modułowy system planowania zasobów przedsiębiorstwa wykorzystywany do zarządzania i optymalizacji różnych aspektów procesu budowlanego.

Cechy - w uczeniu maszynowym niezależne zmienne lub atrybuty używane jako dane wejściowe do modelu.

Fizyczny model danych - szczegółowa reprezentacja struktury bazy danych, w tym tabel, kolumn, typów danych, kluczy i indeksów, zoptymalizowana pod kątem określonego systemu DBMS.

FPDF to biblioteka Pythona służąca do tworzenia dokumentów PDF.

Geometric Core to komponent oprogramowania, który zapewnia podstawowe algorytmy do tworzenia, edycji i analizy obiektów geometrycznych w CAD, BIM i innych aplikacjach inżynieryjnych.

HiPPO (Highest Paid Person's Opinion) - podejście do podejmowania decyzji oparte na opinii najwyżej opłacanej osoby w organizacji, a nie na obiektywnych danych.

IDE (Integrated Development Environment) - zintegrowane środowisko programistyczne, kompleksowe narzędzie do pisania, testowania i debugowania kodu (np. PyCharm, VS Code, Jupyter Notebook).

IDS (Information Delivery Specification) to specyfikacja dostarczania informacji, która definiuje wymagania dotyczące danych na różnych etapach projektu.

IFC (Industry Foundation Classes) to format wymiany danych BIM, który zapewnia interoperacyjność między różnymi rozwiązaniami programowymi.

Przemysł 4.0 to koncepcja rozwoju przemysłowego, która łączy możliwości cyfryzacji, automatyzacji i sztucznej inteligencji z potencjałem ludzkim i zrównoważeniem środowiskowym.

Integracja danych to proces łączenia danych z różnych źródeł w jeden, spójny system w celu zapewnienia jednolitego widoku informacji.

Silosy informacyjne to odizolowane systemy przechowywania danych, które nie udostępniają informacji innym systemom, tworząc bariery dla efektywnego wykorzystania danych.

IoT (Internet of Things) to koncepcja łączenia obiektów fizycznych z Internetem w celu gromadzenia, przetwarzania i przesyłania danych.

k-NN (k-Nearest Neighbors) to algorytm uczenia maszynowego, który klasyfikuje obiekty na podstawie podobieństwa do najbliższych sąsiadów w próbce szkoleniowej.

Kaggle to platforma do analizy danych i konkursów uczenia maszynowego.

Kalkulacja - obliczenie kosztu robót budowlanych lub procesów dla określonej jednostki miary (np. 1 m² ściany gipsowo-kartonowej, 1 m³ betonu).

KPI (Key Performance Indicators) to kluczowe wskaźniki wydajności, wymierne metryki wykorzystywane do oceny sukcesu firmy lub konkretnego projektu.

Etykiety - w uczeniu maszynowym docelowe zmienne lub atrybuty, które model powinien przewidywać.

Algorytm uczenia się - proces znajdowania najlepszej hipotezy w modelu odpowiadającej funkcji docelowej przy użyciu zestawu danych szkoleniowych.

Regresja liniowa - statystyczna metoda modelowania związku między zmienną zależną a jedną lub większą liczbą zmiennych niezależnych.

LLM (Large Language Model) - Duży Model Językowy, sztuczna inteligencja przeszkolona do rozumienia i generowania tekstu z ogromnych ilości danych, zdolna do analizowania kontekstu i pisania kodu programowania.

LOD (Level of Detail/Development) - poziom szczegółowości modelu, który określa stopień dokładności geometrycznej i zawartości informacyjnej.

Logiczny model danych to szczegółowy opis encji, atrybutów, kluczy i relacji odzwierciedlających informacje biznesowe i reguły, stanowiący etap pośredni między modelem koncepcyjnym a fizycznym.

Machine Learning (uczenie maszynowe) - klasa technik sztucznej inteligencji, które umożliwiają systemom komputerowym uczenie się i przewidywanie na podstawie danych bez wyraźnego programowania.

Masterformat to system klasyfikacji pierwszej generacji służący do podziału specyfikacji budowlanych na sekcje i dyscypliny.

MEP (Mechanical, Electrical, Plumbing) - systemy inżynierii budowlanej obejmujące elementy mechaniczne, elektryczne i hydrauliczne.

Mesh to siatkowa reprezentacja obiektów 3D składająca się z wierzchołków, krawędzi i ścian.

Model - w uczeniu maszynowym zestaw różnych hipotez, z których jedna przybliża funkcję docelową, która ma być przewidywana lub przybliżana.

Modelowanie danych to proces tworzenia ustrukturyzowanej reprezentacji danych i ich relacji w celu wdrożenia w systemach informatycznych, w tym na poziomie koncepcyjnym, logicznym i fizycznym.

n8n to narzędzie typu open source do automatyzacji przepływu pracy i integracji aplikacji za pomocą podejścia niskokodowego.

Normalizacja - w uczeniu maszynowym proces sprowadzania różnych danych liczbowych do wspólnej skali w celu ułatwienia przetwarzania i analizy.

Inżynieria odwrotna - proces badania urządzenia, funkcjonowania i technologii produkcji obiektu poprzez analizę jego struktury, funkcji i działania. W kontekście danych - wyodrębnianie informacji z zastrzeżonych formatów do wykorzystania w systemach otwartych.

OCR (Optical Character Recognition) to technologia optycznego rozpoznawania znaków, która konwertuje obrazy tekstowe (zeskanowane dokumenty, zdjęcia) do formatu tekstu nadającego się do odczytu maszynowego.

OmniClass to międzynarodowy standard klasyfikacji drugiej generacji do zarządzania informacjami budowlanymi.

Ontologia - system wzajemnych powiązań pojęć, który formalizuje określoną dziedzinę wiedzy.

Open Source - model rozwoju i dystrybucji otwartego oprogramowania, które jest dostępne do swobodnego użytku, nauki i modyfikacji.

Open BIM to koncepcja otwartego BIM, która obejmuje wykorzystanie otwartych standardów i formatów wymiany danych między różnymi rozwiązaniami programowymi.

Otwarte standardy - publicznie dostępne specyfikacje dotyczące realizacji określonego zadania, które umożliwiają różnym systemom interakcję i wymianę danych.

Pandas to biblioteka Pythona typu open source do przetwarzania i analizy danych, zapewniająca struktury danych DataFrame i Series do wydajnej obsługi informacji tabelarycznych.

Paradygmat otwartych danych to podejście do przetwarzania danych, w którym informacje są swobodnie udostępniane do użytku, ponownego wykorzystania i rozpowszechniania przez każdego.

Metoda **parametryczna** to metoda szacowania projektu budowlanego, która wykorzystuje modele statystyczne do szacowania wartości na podstawie parametrów projektu.

PIMS (Project Information Model) to cyfrowy system przeznaczony do organizowania, przechowywania i udostępniania wszystkich informacji o projekcie.

Pipeline - sekwencja procesów przetwarzania danych, od ekstrakcji i transformacji do analizy i wizualizacji.

PMIS (Project Information Management System) to system zarządzania projektami przeznaczony do szczegółowej kontroli zadań na poziomie indywidualnego projektu budowlanego.

Analityka **predykcyjna** to sekcja analityki, która wykorzystuje metody statystyczne i uczenie maszynowe do przewidywania przyszłych wyników na podstawie danych historycznych.

Analityka **preskrytywna** - sekcja analityki, która nie tylko przewiduje przyszłe wyniki, ale także sugeruje optymalne działania w celu osiągnięcia pożądanych rezultatów.

Zastrzeżone formaty to zamknięte formaty danych kontrolowane przez konkretną firmę, które ograniczają możliwość udostępniania informacji i zwiększą zależność od określonego oprogramowania.

QTO (Quantity Take-Off) to proces wyodrębniania ilości elementów z dokumentacji projektowej w celu obliczenia ilości materiałów wymaganych do realizacji projektu.

System zarządzania jakością - system zarządzania jakością, który zapewnia, że procesy i wyniki spełniają ustalone wymagania.

RAG (Retrieval-Augmented Generation) to metoda, która łączy generatywne możliwości modeli językowych z wydobywaniem istotnych informacji z korporacyjnych baz danych, poprawiając dokładność i trafność odpowiedzi.

RDBMS (Relational Database Management System) to system zarządzania relacyjnymi bazami danych, który organizuje informacje w postaci powiązanych ze sobą tabel.

RegEx (wyrażenia regularne) to sformalizowany język do wyszukiwania i przetwarzania ciągów znaków, umożliwiający określanie szablonów do sprawdzania danych tekstowych pod kątem zgodności z określonymi kryteriami.

Regresja to statystyczna metoda analizy zależności między zmiennymi.

Obliczenia CO₂ to metoda szacowania emisji dwutlenku węgla związanej z produkcją i wykorzystaniem materiałów i procesów budowlanych.

Metoda zasobowa to metoda dokonywania szacunków w oparciu o szczegółową analizę wszystkich zasobów (materiałów, robocizny, sprzętu) wymaganych do wykonania prac budowlanych.

RFID (Radio Frequency Identification) to technologia automatycznej identyfikacji obiektów za pomocą sygnałów radiowych, wykorzystywana do śledzenia materiałów, maszyn i personelu.

ROI (Return on Investment) to wskaźnik odzwierciedlający stosunek zysku do zainwestowanych środków, wykorzystywany do oceny efektywności inwestycji.

SaaS (Software as a Service) to model oprogramowania jako usługi, w którym aplikacje są hostowane przez dostawcę i udostępniane użytkownikom przez Internet.

SCM (Supply Chain Management) - zarządzanie łańcuchem dostaw, które obejmuje koordynację i optymalizację wszystkich procesów od zakupu materiałów po dostawę gotowych produktów.

Silosy danych to odizolowane magazyny informacji w organizacji, które nie są zintegrowane z innymi systemami, co utrudnia udostępnianie danych i jest nieefektywne.

SQL (Structured Query Language) to ustrukturyzowany język zapytań używany do pracy z relacyjnymi bazami danych.

SQLite to lekki, wbudowany, wieloplatformowy DBMS, który nie wymaga oddzielnego serwera i obsługuje podstawowe funkcje SQL, szeroko stosowane w aplikacjach mobilnych i systemach wbudowanych.

Dane ustrukturyzowane - informacje zorganizowane w określonym formacie o wyraźnej strukturze, np. w relacyjnych bazach danych lub tabelach.

Dane o luźnej strukturze - informacje o częściowej organizacji i elastycznej strukturze, takie jak JSON lub XML, w których różne elementy mogą zawierać różne zestawy atrybutów.

Encja to konkretny lub abstrakcyjny obiekt świata rzeczywistego, który można jednoznacznie zidentyfikować, opisać i przedstawić w postaci danych.

Uczenie nadzorowane - rodzaj uczenia maszynowego, w którym algorytm jest szkolony na podzielonych danych, gdzie pożądany wynik jest znany dla każdego przykładu.

Taksonomia to hierarchiczny system klasyfikacji używany do systematycznego kategoryzowania elementów w oparciu o wspólne cechy.

Titanic Dataset to popularny zbiór danych do trenowania i testowania modeli uczenia maszynowego.

Trening - proces, w którym algorytm uczenia maszynowego analizuje dane w celu zidentyfikowania wzorców i utworzenia modelu.

Transfer learning to technika uczenia maszynowego, w której model wytrenowany dla jednego zadania jest wykorzystywany jako punkt wyjścia dla innego zadania.

Transformacja - proces zmiany formatu, struktury lub zawartości danych w celu ich późniejszego wykorzystania.

Wymagania dotyczące danych - sformalizowane kryteria określające strukturę, format, kompletność i jakość informacji wymaganych do obsługi procesów biznesowych.

Uberyzacja branży budowlanej to proces transformacji tradycyjnych modeli biznesowych w budownictwie pod wpływem platform cyfrowych, które zapewniają bezpośrednią interakcję między klientami a wykonawcami bez pośredników.

Uniclass to system klasyfikacji elementów budowlanych drugiej i trzeciej generacji szeroko stosowany w Wielkiej Brytanii.

USD (Universal Scene Description) to format danych opracowany dla grafiki komputerowej, ale zyskał zastosowanie w systemach inżynierijnych ze względu na prostą strukturę i niezależność od rdzeni geometrycznych.

Walidacja danych to proces sprawdzania informacji pod kątem ustalonych kryteriów i wymogów w celu zapewnienia dokładności, kompletności i spójności danych.

Wektorowa baza danych - wyspecjalizowany typ bazy danych, który przechowuje dane jako wielowymiarowe wektory w celu wydajnego wyszukiwania semantycznego i porównywania obiektów.

Reprezentacja wektorowa (embedding) to metoda przekształcania danych w wielowymiarowe wektory numeryczne, która umożliwia algorytmom maszynowym wydajne przetwarzanie i analizowanie informacji.

VectorOps to metodologia skoncentrowana na przetwarzaniu, przechowywaniu i analizowaniu wielowymiarowych danych wektorowych, szczególnie istotnych w obszarach takich jak cyfrowe bliźniaki i wyszukiwanie semantyczne.

Wizualizacja - graficzna reprezentacja danych w celu lepszego postrzegania i analizy informacji.

Alfabetyczna kategoryzacja terminów została przeprowadzona według ich angielskich nazw.

LISTA REFERENCJI I MATERIAŁÓW ONLINE

- [1] Gartner, "IT Key Metrics Data 2017: Index of Published Documents and Metrics", 12 grudnia 2016 r. [Online]. Dostępne: <https://www.gartner.com/en/documents/3530919>. [Data dostępu: 1 marca 2025 r.].
- [2] KPMG, "Znane wyzwania - nowe podejścia. 2023 Global Construction Survey", 1 stycznia 2023 r. [Online]. Dostępne: <https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2023/06/familiar-challenges-new-solutions-1.pdf>. [Dostęp: 5 marca 2025 r.].
- [3] F. R. Barnard, "A picture is worth a thousand words", 10 Mapi 1927. [Online]. Dostępne: https://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words. [Data odniesienia: 15 marca 2025 r.].
- [4] M. Bastian, "Microsoft CEO Satya Nadella says self-claiming AGI is 'nonsensical benchmark hacking', 21 Feb. 2025. [Online]. Dostępne: <https://the-decoder.com/microsoft-ceo-satya-nadella-says-self-claimingagi-is-nonsensical-benchmark-hacking/>. [Data dostępu: 15 marca 2025 r.].
- [5] W. E. Forum, "Forum Shaping the Future of Construction - A Landscape in Transformation:", 1 stycznia 2016 r. [Online]. Dostępne: https://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Construction.pdf. [Data dostępu: 2 marca 2025 r.].
- [6] C. D. Gillespie, "Clay: Earth's entanglement in the age of clay", 2024. [Online]. Dostępne: <https://ufl.pb.unizin.org/imos/chapter/clay/>.
- [7] "Papirus z III wieku p.n.e. Język to grecki", 2024. [Online]. Available: <https://www.facebook.com/429710190886668/posts/595698270954525>.
- [8] "Monitorowanie: korzystanie z dostępnych narzędzi", 1980. [Online]. Dostępne: <https://pubmed.ncbi.nlm.nih.gov/10246720/>. [Data adresu: 15 marca 2025 r.].
- [9] PWC, "Data driven What students need to succeed in a rapidly changing business world", 15 lutego 2015 r. [Online]. Dostępne: <https://www.pwc.com/us/en/faculty-resource/assets/PwC-Data-driven-paper-Feb2015.pdf>. [Data dostępu: 15 marca 2025 r.].
- [10] Skanska USA, "Jesienne trendy na rynku budowlanym", 2 listopada 2023 r. [Online]. Available: <https://x.com/SkanskaUSA/status/1720167220817588714>.
- [11] "Oxford Essential Quotations (4th ed.)", Oxford University Press, 2016. [Online]. Available: <https://www.oxfordreference.com/display/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00006236>. [Data dostępu: 1 marca 2025].

- [12] "Cytat: Sondergaard o analityce danych," [Online]. Dostępne: <https://www.causeweb.org/cause/resources/library/r2493>. [Data dostępu: 15 marca 2025 r.].
- [13] "Jak globalne zainteresowanie sztuczną inteligencją pobudza rynek zarządzania danymi", 28 maja 2024 r. [Online]. Dostępne: <https://iot-analytics.com/how-global-ai-interest-is-boosting-data-management-market/>. [Data dostępu: 15 marca 2025 r.].
- [14] И. McCue, "Historia ERP", 2024. [Online]. Dostępne: <https://www.netsuite.com/portal/resource/articles/erp/erp-history.shtml>.
- [15] erpscout, "Cena ERP: Ile kosztuje system ERP?," [Online]. Dostępny: <https://erpscout.de/en/erp-costs/>. [Data dostępu: 15 marca 2025 r.].
- [16] softwarepath, "Co 1 384 projekty ERP mówią nam o wyborze ERP (raport ERP 2022)", 18 stycznia 2022. [Online]. Dostępne: <https://softwarepath.com/guides/erp-report>. [Data dostępu: 15 marca 2025 r.].
- [17] Deloitte, "Zarządzanie oparte na danych w cyfrowych projektach kapitałowych", 16 grudnia 2016 r. [Online]. Dostępne: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/Real%20Estate/us-engineering-construction-data-driven-management-digital-capital-projects.pdf>. [Data adresu: 1 marca 2025 r.].
- [18] Mckinsey, "The data-driven enterprise of 2025", 28 stycznia 2022 r. [Online]. Dostępne: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025>. [Data odniesienia: 22 maja 2024 r.].
- [19] Wikipedia, "Prawo Moore'a," [Online]. Dostępny: https://en.wikipedia.org/wiki/Moore%27s_law. [Data dostępu: 15 marca 2025 r.].
- [20] Accenture, "Budowanie większej wartości dzięki projektom kapitałowym", 1 stycznia 2020 r. [Online]. Dostępne: <https://www.accenture.com/content/dam/accenture/final/a-com-migration/r3-3/pdf/pdf-143/accenture-industryx-building-value-capital-projects-highres.pdf>. [Data dostępu: 3 marca 2024 r.].
- [21] B. Marr, "Ile danych tworzymy każdego dnia? Zaskakujące statystyki, które każdy powinien przeczytać", 2018. [Online]. Dostępne: <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read>.
- [22] "Ile danych jest produkowanych każdego dnia?", 2024. [Online]. Dostępne: <https://graduate.northeastern.edu/resources/how-much-data-produced-every-day/>.
- [23] T. Sullivan, "AI and the global 'datasphere': how much information will humanity have by 2025?", 2024. [Online]. Dostępne: <https://www.datauniverseevent.com/en-us/blog/general/AI-and-the-Global-Datasphere-How-Much-Information-Will-Humanity-Have-By-2025.html>.

- [24] Statista, "Łączna liczba wydrukowanych książek wyprodukowanych w różnych regionach Europy Zachodniej w każdym półwieczu w latach 1454-1800", [Internet]. [Online]. Dostępne: <https://www.statista.com/statistics/1396121/europe-book-production-half-century-region-historical/>. [Data odniesienia: 1 marca 2025 r.].
- [25] "Przykłady cen", 2024. [Online]. Dostępne: <https://cloud.google.com/storage/pricing-examples>.
- [26] M. Ashare, "Enterprises outsourcing data storage as complexity rises", 10 maja 2024. [Online]. Dostępne: <https://www.ciodive.com/news/enterprises-outsource-data-storage-complexity-rises/715854/>. [Data dostępu: 15 marca 2025 r.].
- [27] JETSOFTPRO, "SaaS is Dead? Szokująca prognoza prezesa Microsoftu wyjaśniona", 13 stycznia 2025 r. [Online]. Dostępne: <https://jetsoftpro.com/blog/saas-is-dead/>.
- [28] BG2 Pod, "Satya Nadella | BG2 w/ Bill Gurley & Brad Gerstner," 12 grudnia 2024. [Online]. Dostępne: https://www.youtube.com/watch?v=9NtsnzRFJ_o. [Data dostępu: 15 marca 2025 r.].
- [29] GoodReads, "Tim Berners-Lee," [Online]. Dostępny: <https://www.goodreads.com/quotes/8644920-data-is-a-precious-thing-and-will-last-longer-than>. [Data dostępu: 15 marca 2025 r.].
- [30] KPMG, "Cue Construction 4.0: Make-or-Break Time", 1 stycznia 2023 r. [Online]. Dostępne: <https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [Dostęp: 5 marca 2025 r.].
- [31] I. Deininger, B. Koch, R. Bauknecht i M. Langhans, "Wykorzystanie modeli cyfrowych do dekarbonizacji zakładu produkcyjnego: przykład połączenia modelu budynku, modelu produkcji i modelu energetycznego", 2024. [Online]. Dostępne: https://www.researchgate.net/publication/374023998_Using_Digital_Models_to_Decarbonize_a_Production_Site_A_Case_Study_of_Connecting_the_Building_Model_Production_Model_and_Energy_Model.
- [32] McKinsey, "REINVENTING CONSTRUCTION: A ROUTE TO HIGHER PRODUCTIVITY", 1 lutego 2017 r. [Online]. Dostępne: <https://www.mckinsey.com/~/media/mckinsey/business%20functions/operations/our%20insights/reinventing%20construction%20through%20a%20productivity%20revolution/mgi-reinventing-construction-a-route-to-higher-productivity-full-report.pdf>.
- [33] Grupa zadaniowa ds. budownictwa do wicepremiera, "Rethinking Construction", 1 października 2014 r. [Online]. Dostępne: https://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking_construction_report.pdf.
- [34] Forbes, "Without An Opinion, You're Just Another Person With Data," 15 Mar. 2016. [Online]. Dostępne: <https://www.forbes.com/sites/silberzahnjones/2016/03/15/without-an-opinion-youre-just-another-person-with-data/>. [Data dostępu: 15 marca 2025 r.].

- [35] Wikiquote, "Charles Babbage," [Online]. Dostępny: https://en.wikiquote.org/wiki/Charles_Babbage. [Data dostępu: 15 marca 2025 r.].
- [36] SAP, "Nowe badania wykazują, że prawie połowa kadry kierowniczej ufa sztucznej inteligencji bardziej niż sobie", 12 marca 2025 r. [Online]. Dostępne: <https://news.sap.com/2025/03/new-research-executive-trust-ai/>. [Data dostępu: 15 marca 2025 r.].
- [37] Kanadyjskie Stowarzyszenie Budownictwa i KPMG w Kanadzie, 2021 r., "Budownictwo w cyfrowym świecie", 1 maja 2021 r. [Online]. Dostępne: <https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf>. [Data dostępu: 5 marca 2025 r.].
- [38] ZCS, "Decoding the Fifth Industrial Revolution," [Online]. Dostępny: <https://www.pwc.in/decoding-the-fifth-industrial-revolution.html>. [Data dostępu: 15 marca 2025 r.].
- [39] M. K., Private Rights and Public Problems: The Global Economics of, Peterson Inst. for Intern. Economics, 2012.
- [40] F. N. a. Y. Z. Harvard Business School: Manuel Hoffmann, "The Value of Open Source Software", 24 stycznia 2024. [Online]. Dostępne: <https://www.hbs.edu/faculty/Pages/item.aspx?num=65230>. [Data dostępu: 15 marca 2025 r.].
- [41] Naval Centre for Cost Analysis Air Force Cost Analysis Agency, "Software Development Cost Estimating Handbook", 1 września 2008 r. [Online]. Dostępne: <https://www.dau.edu/sites/default/files/Migrated/CopDocuments/SW%20Cost%20Est%20Manual%20Vol%20I%20rev%2010.pdf>.
- [42] McKinsey, "Poprawa wydajności w budownictwie", [Online]. Dostępne: <https://www.mckinsey.com/capabilities/operations/our-insights/improving-construction-productivity>. [Data adresu: 15 marca 2025 r.].
- [43] A. G. a. C. Syverson, "The Strange and Awful Path of Productivity in the US Construction Sector", 19 stycznia 2023. [Online]. Dostępne: <https://bfi.uchicago.edu/insight/research-summary/the-strange-and-awful-path-of-productivity-in-the-us-construction-sector/>. [Data dostępu: 1 marca 2025 r.].
- [44] McKinsey, "Zapewnienie wydajności w budownictwie nie jest już opcjonalne", 9 sierpnia 2024 r. [Online]. Dostępne: <https://www.mckinsey.com/capabilities/operations/our-insights/delivering-on-construction-productivity-is-no-longer-optional>. [Data dostępu: 5 marca 2025 r.].
- [45] Grupa ING, "Opóźniająca się produktywność w budownictwie powoduje wzrost kosztów budowy", 12 grudnia 2022 r. [Online]. Dostępne: <https://think.ing.com/articles/lagging-productivity-drives-up-building-costs-in-many-eu-countries/>. [Data publikacji: 15 marca 2025 r.].
- [46] M. Berman, "Microsoft CEO's Shocking Prediction: 'Agents Will Replace ALL Software'", 19 grudnia 2024. [Online]. Dostępne: <https://www.youtube.com/watch?v=uGOLYz2pgr8>. [Data dostępu: 15

- marca 2025 r.].
- [47] Business Insider, "Dyrektor generalny Anthropic twierdzi, że w ciągu 3 do 6 miesięcy sztuczna inteligencja będzie pisać 90% kodu, za który odpowiedzialni byli programiści", 15 marca 2025 r. [Online]. Dostępne: <https://www.businessinsider.com/anthropic-ceo-ai-90-percent-code-3-to-6-months-2025-3>. [Data dostępu: 30 marca 2025 r.].
- [48] Statista, "Porównanie popularności systemów zarządzania bazami danych (DBMS) na całym świecie w czerwcu 2024 r. według kategorii", czerwiec 2024 r. [Online]. Dostępne: <https://www.statista.com/statistics/1131595/worldwide-popularity-database-management-systems-category/>. [Data adresu: 15 marca 2025 r.].
- [49] DB-Engines, "Ranking DB-Engines", [Online]. Dostępne: <https://db-engines.com/en/ranking>. [Data odniesienia: 15 marca 2025 r.].
- [50] "Stack Overflow Developer Survey 2023", 2024. [Online]. Dostępne: <https://survey.stackoverflow.co/2023/>.
- [51] "SQL", 2024. [Online]. Dostępne: <https://en.wikipedia.org/wiki/SQL>.
- [52] "Dane ustrukturyzowane i nieustrukturyzowane: jaką jest różnica?", 2024. [Online]. Dostępne: <https://www.ibm.com/blog/structured-vs-unstructured-data/>.
- [53] DataDrivenConstruction, "PDF PORÓWNANIE FORMATÓW DANYCH DLA PROJEKTÓW BUDOWLANYCH", 23 kwietnia 2024. [Online]. Dostępne: <https://datadrivenconstruction.io/wp-content/uploads/2024/10/COMPARISON-OF-DATA-FORMATS-FOR-CONSTRUCTION-PROJECTS-1.pdf>.
- [54] "Building Information Modeling Whitepaper site", 2003. [Online]. Dostępne: <https://web.archive.org/web/20030711125527/http:/usa.autodesk.com/adsk/servlet/item?id=2255342&siteID=123112>.
- [55] A. Boyko, "Wojny lobbingowe i rozwój BIM. Część 5: BlackRock jest mistrzem wszystkich technologii. Jak korporacje kontrolują otwarty kod źródłowy", 2024. [Online]. Dostępne: <https://bigdataconstruction.com/autodesk-oracle-blackrock-open-source/>.
- [56] D. Ushakov, "Direct Modeling - Who and Why Needs It? A Review of Competitive Technologies," 14 listopada 2011. [Online]. Dostępne: https://isicad.net/articles.php?article_num=14805. [Data adresu: 02 2025].
- [57] C. Eastman i A. Cthers, "Eastman, Charles; And Cthers", wrzesień 1974. [Online]. Dostępny: <https://files.eric.ed.gov/fulltext/ED113833.pdf>. [Data dostępu: 15 marca 2025 r.].
- [58] D. Ushakov, "Direct Modeling - Who and Why Needs It? A Review of Competitive Technologies," 11 listopada 2011. [Online]. Dostępne: https://isicad.net/articles.php?article_num=14805. [Data odniesienia: 15 marca 2025 r.].

- [59] D. Weisberg, "Historia CAD", 12 grudnia 2022 r. [Online]. Dostępne: https://www.shapr3d.com/blog/history-of-cad?utm_campaign=cadhistorynet. [Data odniesienia: 15 marca 2025 r.].
- [60] ADSK, "Biała Księga Modelowania Informacji o Budynku", 2002. [Online]. Dostępne: https://web.archive.org/web/20060512180953/http://images.adsk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Data odniesienia: 15 marca 2025 r.].
- [61] ADSK, "Biała księga Modelowanie informacji o budynku w praktyce", [Online]. Dostępny: https://web.archive.org/web/20060512181000/http://images.adsk.com/apac_sapac_main/files/4525077_BIM_in_Practice.pdf. [Data dostępu: 15 marca 2025 r.].
- [62] A. Boyko, "Wojny lobbingowe i rozwój BIM. Część 2: otwarty BIM VS zamknięty BIM. Europa VS reszta świata", 2024. [Online]. Dostępne: <https://bigdataconstruction.com/lobbyist-wars-and-the-development-of-bim-part-2-open-bim-vs-closed-bim-revit-vs-archicad-and-europe-vs-the-rest-of-the-world/>.
- [63] A. Boyko, "Lobbykriege um Daten im Bauwesen | Techno-Feudalismus und die Geschichte von BIMs", 2024. [Online]. Dostępne: https://youtu.be/S-TNdUgfHxk?si=evM_v28KQbGOG0k&t=1360.
- [64] ADSK, "Whitepaper BIM", 2002. [Online]. Dostępne: https://web.archive.org/web/20060512180953/http://images.autodesk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Data adresu: 15 marca 2025 r.].
- [65] ADSK, "Integrated Design-Through-Manufacturing: Benefits and Rationale," [Online]. Dostępne: https://web.archive.org/web/20010615093351/http://www3.adsk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Data dostępu: 15 marca 2025 r.].
- [66] M. Shacklett, "Dane ustrukturyzowane i nieustrukturyzowane: Kluczowe różnice", 2024. [Online]. Dostępne: <https://www.datamation.com/big-data/structured-vs-unstructured-data/>.
- [67] K. Woolard, "Making sense of the growth of unstructured data", 2024. [Online]. Dostępne: <https://automationhero.ai/blog/making-sense-of-the-rise-of-unstructured-data/>.
- [68] A. C. O. J. L. D. J. a. L. T. G. Michael P. Gallaher, "Cost Analysis of Inadequate Interoperability in the," 2004. [Online]. Dostępny: <https://nvlpubs.nist.gov/nistpubs/gcr/2004/nist.gcr.04-867.pdf>. [Data dostępu: 02 2025].
- [69] CrowdFlower, "Data Science Report 2016", 2016. [Online]. Dostępne: https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf. [Data dostępu: 15 marca 2025 r.].
- [70] Analyticsindiamag, "6 najbardziej czasochłonnych zadań dla analityków danych", 15 maja 2019 r. [Online]. Dostępne: <https://analyticsindiamag.com/ai-trends/6-tasks-data-scientists-spend-the-most-time-doing/>.

- [71] BizReport, "Report: Data scientists spend bulk of time cleaning up," 06 July 2015. [Online]. Dostępne: <https://web.archive.org/web/20200824174530/http://www.bizreport.com/2015/07/report-data-scientists-spend-bulk-of-time-cleaning-up.html>. [Data dostępu: 5 marca 2025 r.].
- [72] S. Hawking, "Science AMA Series: Stephen Hawking AMA Answers!", 27 lipca 2015 r. [Online]. Dostępne: https://www.reddit.com/r/science/comments/3nyn5i/science_ama_series_stephen_hawking_ama_answers/. [Data dostępu: 15 marca 2025 r.].
- [73] B. Cyphers i K. Doctorow, "Prywatność bez monopolu: ochrona danych i interoperacyjność", 2024. [Online]. Dostępne: <https://www.eff.org/wp/interoperability-and-privacy>.
- [74] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information", 1 października 2013 r. [Online]. Dostępne: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>. [Data dostępu: 15 marca 2025 r.].
- [75] A. Boyko, "Walka o otwarte dane w branży budowlanej. Historia AUTOLISP, intelliCAD, openDWG, ODA i openCASCADE", 15 05 2024. [Online]. Dostępne: <https://boikoartem.medium.com/the-struggle-for-open-data-in-the-construction-industry-2b97200e6393>. [Data adresu: 16 02 2025].
- [76] Wikipedia, "Microsoft i open source", [Online]. Dostępny: https://en.wikipedia.org/wiki/Microsoft_and_open_source. [Data dostępu: 15 marca 2025 r.].
- [77] TIME, "Przepaść między otwartymi i zamkniętymi modelami sztucznej inteligencji może się zmniejszać. Oto dlaczego to ma znaczenie", 5 listopada 2024 r. [Online]. Dostępne: <https://time.com/7171962/open-closed-ai-models-epoch/>. [Data dostępu: 15 marca 2025 r.].
- [78] The Verge, "Ponad jedna czwarta nowego kodu w Google jest generowana przez sztuczną inteligencję", 29 października 2024 r. [Online]. Available: <https://www.theverge.com/2024/10/29/24282757/google-new-code-generated-ai-q3-2024>. [Data dostępu: 15 marca 2025 r.].
- [79] McKinsey Digital, "The business case for using GPUs to accelerate analytics processing", 15 grudnia 2020 r. [Online]. Dostępne: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/the-business-case-for-using-gpus-to-accelerate-analytics-processing>. [Dostęp: 15 marca 2025 r.].
- [80] PwC, "PwC Open Source Monitor 2019", 2019. [Online]. Dostępne: <https://www.pwc.de/de/digitale-transformation/open-source-monitor-research-report-2019.pdf>. [Dostęp 15 marca 2025 r.].
- [81] Travers Smith, "The Open Secret: Open Source Software", 2024. [Online]. Dostępny: <https://www.traverssmith.com/knowledge/knowledge-container/the-open-secret-open-source-software/>. [Data dostępu: 15 marca 2025 r.].

- [82] Deloitte, "Proces transferu danych w transformacjach korporacyjnych", 2021. [Online]. Dostępne: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/us-the-data-transfer-process-in-corporate-transformations.pdf>. [Data odniesienia: 15 marca 2025 r.].
- [83] gov.uk, "Data Analytics and AI in Government Project Delivery", 20 marca 2024 r. [Online]. Dostępne: <https://www.gov.uk/government/publications/data-analytics-and-ai-in-government-project-delivery/data-analytics-and-ai-in-government-project-delivery>. [Data dostępu: 5 marca 2025 r.].
- [84] "Cytat Origin: Wszystko powinno być tak proste, jak to tylko możliwe, ale nie prostsze", 13 maja 2011. [Online]. Dostępny: <https://quoteinvestigator.com/2011/05/13/einstein-simple/>. [Data dostępu: 15 marca 2025].
- [85] "Transformer (architektura głębokiego uczenia)," [Online]. Dostępne: [https://en.wikipedia.org/wiki/Transformer_\(deep_learning_architecture\)](https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture)). [Data dostępu: 15 marca 2025 r.].
- [86] "Statystyki pobierania pakietów Python", 2024. [Online]. Dostępne: <https://www.pepy.tech/projects/pandas>.
- [87] Interview Bit, "10 najlepszych bibliotek Pythona", 2023. [Online]. Dostępne: <https://www.interviewbit.com/blog/python-libraries/#:~:text=With%20more%20than%20137%2C000%20libraries,data%20manipulation%2C%20and%20many%20more>. [Data dostępu: 30 marca 2025 r.].
- [88] "NVIDIA i HP Supercharge Data Science and Generative AI on Workstations", 7 marca 2025 r. [Online]. Dostępne: <https://nvidianews.nvidia.com/news/nvidia-hp-supercharge-data-science-generative-ai-workstations>. [Data dostępu: 15 marca 2025 r.].
- [89] P. Orac, "How to process a DataFrame with millions of rows in seconds", 2024. [Online]. Dostępne: <https://towardsdatascience.com/how-to-process-a-dataframe-with-millions-of-rows-in-seconds>.
- [90] Ç. Uslu, "Co to jest Kaggle?", 2024. [Online]. Dostępne: <https://www.datacamp.com/blog/what-is-kaggle>.
- [91] "NVIDIA CEO Jensen Huang Keynote na COMPUTEX 2024", 2 czerwca 2024 r. [Online]. Dostępne: <https://www.youtube.com/live/pKXDVsWZmUU?si=Z3Rj1Las8wiPII2w>. [Dostęp 15 marca 2025 r.].
- [92] "Członkowie: założyciele i członkowie korporacyjni", 2024. [Online]. Dostępne: <https://www.opendesign.com/member-showcase>.
- [93] A. Boyko, "The Age of Change: IFC is a thing of the past or why ADSK and other CAD vendors are willing to give up IFC for USD in 14 key facts", 24 listopada 2024. [Online]. Dostępne: <https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-are-willing-to-3f9a82cccd10a>. [Data dostępu: 23 lutego 2025 r.].

- [94] A. Boyko, "Świat post-BIM. Przejście na dane i procesy oraz czy branża budowlana potrzebuje semantyki, formatów i interoperacyjności", 20 grudnia 2024. [Online]. Dostępne: <https://boikoartem.medium.com/the-post-bim-world-7e35b7271119>. [Data adresu: 23 lutego 2025 r.].
- [95] N. I. o. Health, "NIH STRATEGIC PLAN FOR DATA SCIENCE", 2016. [Online]. Dostępne: https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf. [Data dostępu: 23 lutego 2025 r.].
- [96] Harvard Business Review, "Bad Data Costs the U.S. \$3 Trillion Per Year", 22 września 2016 r. [Online]. Dostępne: <https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year>.
- [97] Delpha, "Wpływ jakości danych", 1 stycznia 2025 r. [Online]. Dostępne: <https://delpha.io/impacts-of-data-quality/>.
- [98] W. B. D. Guide, "Design for Maintainability: The Importance of Operations and Maintenance Considerations During the Design Phase of Construction Projects," [Online]. Dostępne: <https://www.wbdg.org/resources/design-for-maintainability>. [Data dostępu: 15 marca 2025 r.].
- [99] O. o. D. C. P. a. Oversight, "Corrosion Prevention and Control Planning Guidebook for Military Systems and Equipment", kwiecień 2014. [Online]. Dostępne: <https://www.dau.edu/sites/default/files/Migrated/CopDocuments/CPC%20Planning%20Guidebook%204%20Feb%202014.pdf>. [Data dostępu: 15 marca 2025 r.].
- [100] Gartner, "Data Quality: Best Practices for Accurate Insights", 1 stycznia 2025 r. [Online]. Dostępne: <https://www.gartner.com/en/data-analytics/topics/data-quality>.
- [101] "For Want of a Nail," [Online]. Dostępny: https://en.wikipedia.org/wiki/For_Want_of_a_Nail. [Data dostępu: 15 marca 2025 r.].
- [102] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information", październik 2013. [Online]. Dostępny: https://www.mckinsey.com/~media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/open%20data%20unlocking%20innovation%20and%20performance%20with%20liquid%20information/mgi_open_data_fullreport_oct2013.pdf. [Data dostępu: 15 marca 2025 r.].
- [103] EY, "Droga do neutralności węglowej", 10 marca 2023 r. [Online]. Dostępne: https://www.ey.com/ru_kz/services/consulting/the-path-to-carbon-neutrality. [Data dostępu: 15 marca 2025 r.].
- [104] PWC, "ESG Awareness", 1 lipca 2024 r. [Online]. Dostępne: <https://www.pwc.com/kz/en/assets/esg-awareness/kz-esg-awareness-rus.pdf>. [Data dostępu: 15 marca 2025 r.].
- [105] G. Hammond, "Embodied Carbon - The Inventory of Carbon and Energy (ICE)", 2024. [Online]. Dostępne: <https://greenbuildingencyclopaedia.uk/wp-content/uploads/2014/07/Full-BSRIA-ICE->

guide.pdf.

- [106] "CO2_calculating the embodied carbon", 2024. [Online]. Dostępne: https://github.com/datadrivenconstruction/CO2_calculating-the-embodied-carbon.
- [107] McKinsey, "Imagining Construction's Digital Future", 24 czerwca 2016 r. [Online]. Dostępne: <https://www.mckinsey.com/capabilities/operations/our-insights/imagining-constructions-digital-future>. [Data odniesienia: 25 lutego 2025 r.].
- [108] Bund der Steuerzahler Deutschland e.V., "Das Schwarzbuch", 10 października 2024 r. [Online]. Dostępne: <https://steuerzahler.de/aktuelles/detail/das-schwarzbuch-202425/>. [Data adresu: 15 marca 2025 r.].
- [109] SAS, "Jezioro danych i hurtownia danych - poznaj różnicę", [Online]. Dostępne: https://www.sas.com/en_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html. [Data odniesienia: 15 marca 2025 r.].
- [110] ADSK, "Modelowanie informacji o budynku", 2002. [Online]. Dostępne: https://www.laiserin.com/features/bim/autodesk_bim.pdf. [Data adresu: 15 marca 2025 r.].
- [111] A. Boyko, "Mapa historii BIM", 2024. [Online]. Dostępne: <https://bigdataconstruction.com/history-of-bim/>.
- [112] A. S. Borkowski, "Definicje BIM według organizacji i standardów", 27 grudnia 2023 r. [Online]. Dostępne: <https://encyclopedia.pub/entry/53149>. [Data adresu: 5 marca 2025 r.].
- [113] CAD Vendor, "Program OPEN BIM", 2012. [Online]. Dostępne: https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open_bim/. [Data odniesienia: 30 marca 2025 r.].
- [114] Wikipedia, "Industry Foundation Classes," [Online]. Dostępny: https://en.wikipedia.org/wiki/Industry_Foundation_Classes. [Data dostępu: 15 marca 2025 r.].
- [115] Wikipedia, "IGES," [Online]. Dostępny: <https://en.wikipedia.org/wiki/IGES>. [Data odniesienia: 30 marca 2025 r.].
- [116] A. Boyko, "Historia CAD (BIM)", 15 grudnia 2021 r. [Online]. Dostępne: https://miro.com/app/board/o9J_IaML2cs=/. [Data adresu: 24 lutego 2025 r.].
- [117] T. K. K. A. O. F. B. C. E. L. H. H. H. E. L. P. N. S. H. T. J. v. L. H. H. G. D. H. T. K. C. L. A. W. J. S. Francesca Noardo, "Reference study of IFC software support: the GeoBIM benchmark 2019 - Part I", 8 stycznia 2021. [Online]. Dostępne: <https://arxiv.org/pdf/2007.10951.pdf>. [Data dostępu: 5 marca 2025 r.].
- [118] И. Rogachev, "Let's Talk BIM: Maxim Nechiporenko | Renga | IFC | Domestic BIM," 13 kwietnia 2021. [Online]. Dostępne: <https://www.youtube.com/watch?t=3000&v=VO3Y9uuuzF9M&feature=youtu.be>.

- [Data adresu: 5 marca 2025 r.].
- [119] D. Ares, "RETS in Real Estate: Why It's Crucial for Efficiency & Growth", 17 grudnia 2024. [Online]. Dostępne: <https://www.realpha.com/blog/rets-importance-in-real-estate-explained>. [Data dostępu: 5 marca 2025 r.].
- [120] "Koszt tokena Flex", 2024. [Online]. Dostępne: <https://www.adsk.com/buying/flex?term=1-YEAR&tab=flex>.
- [121] A. Boyko, "Zapomnij o BIM i zdemokratyzuj dostęp do danych (17. Kolokwium Inwestor - Hochschule - Bauindustrie)", 2024. [Online]. Dostępne: <https://www.bim.bayern.de/wp-content/uploads/2023/06/Kolloquium-17-TUM-Baprozessmanagment-und-Bay-Bauindustrie.pdf>.
- [122] D. Hill, D. Foldesi, S. Ferrer, M. Friedman, E. Loh i F. Plaschke, "Solving the construction industry productivity puzzle", 2015. [Online]. Dostępne: <https://www.bcg.com/publications/2015/engineered-products-project-business-solving-construction-industrys-productivity-puzzle>.
- [123] "SCOPE - Projektdatenumgebung und Modellierung multifunktionaler Bauprodukte mit Fokus auf die Gebäudehülle," 1 stycznia 2018. [Online]. Dostępne: <https://www.ise.fraunhofer.de/de/forschungsprojekte/scope.html>. [Data dostępu: 2 marca 2025 r.].
- [124] Apple.com, "Pixar, Adobe, Apple i NVIDIA tworzą sojusz na rzecz OpenUSD, aby promować otwarte standardy dla treści 3D", 1 sierpnia 2023 r. [Online]. Dostępne: <https://www.apple.com/newsroom/2023/08/pixar-adobe-apple-adsk-and-nvidia-form-alliance-for-openusd/>. [Dostęp: 2 marca 2025 r.].
- [125] AECmag, "Strategia danych granularnych ADSK", 25 lipca 2024 r. [Online]. Dostępne: <https://aecmag.com/technology/autodesks-granular-data-strategy/>. [Data dostępu: 15 marca 2025 r.].
- [126] A. Boyko, "Era zmian: IFC to już przeszłość, czyli dlaczego ADSK i inni dostawcy CAD są gotowi zrezygnować z IFC na rzecz USD w 14 kluczowych faktach", 24 11 2024. [Online]. Dostępne: <https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-are-willing-to-3f9a82cccd10a>. [Data adresu: 23 lutego 2025 r.].
- [127] A. Boyko, "ENG BIM Cluster 2024 | Bitwa o dane i zastosowanie LLM i ChatGPT w budownictwie", 7 sierpnia 2024. [Online]. Dostępny: ENG BIM Cluster 2024 | Bitwa o dane i zastosowanie LLM i ChatGPT w budownictwie. [Dostęp: 15 marca 2025 r.].
- [128] "Jeffrey Zeldman Presents", 6 maja 2008 r. [Online]. Dostępny: <https://zeldman.com/2008/05/06/content-precedes-design/>. [Data dostępu: 15 marca 2025].
- [129] A. Boyko, "DWG Analyse with ChatGPT | DataDrivenConstruction," 5 Mar. 2024. [Online]. Dostępne: <https://www.kaggle.com/code/artemboiko/dwg-analyse-with-chatgpt-datadrivenconstruction>. [Data odniesienia: 15 marca 2025 r.].

- [130] McKinsey , "The McKinsey guide to outcompeting in the age of digital and AI", 2023. [Online]. Dostępne: <https://www.mckinsey.com/featured-insights/mckinsey-on-books/rewired>. [Data adresu: 30 marca 2025 r.].
- [131] Forbes, "Data Storytelling: The Essential Data Science Skill Everyone Needs", 31 marca 2016 r. [Online]. Dostępne: <https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/>. [Data dostępu: 15 marca 2025 r.].
- [132] J. Bertin, "Grafika i graficzne przetwarzanie informacji", 8 września 2011 r. [Online]. Dostępny: https://books.google.de/books/about/Graphics_and_Graphic_Information_Process.html?id=csqX_xnm4tcC&redir_esc=y. [Data dostępu: 15 marca 2025].
- [133] CauseWeb, "Wells/Wilks on Statistical Thinking" [Online]. Dostępny: <https://www.causeweb.org/cause/resources/library/r1266>. [Data dostępu: 15 marca 2025 r.].
- [134] Ministrymagazine, "Jak nauka odkryła Stworzenie", styczeń 1986. [Online]. Dostępny: <https://www.ministrymagazine.org/archive/1986/01/how-science-discovered-creation>. [Data odniesienia: 15 marca 2025 r.].
- [135] BCG, "Data-Driven Transformation: Accelerate at Scale Now", 23 maja 2017 r. [Online]. Dostępny: <https://www.bcg.com/publications/2017/digital-transformation-transformation-data-driven-transformation>. [Dostęp 15 maja 2024 r.].
- [136] "Jak zbudować architekturę danych, aby napędzać innowacje - dziś i jutro", 3 czerwca 2020 r. [Online]. Dostępne: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-to-build-a-data-architecture-to-drive-innovation-today-and-tomorrow>. [Data adresu: 15 marca 2025 r.].
- [137] Oxford, "Woodrow Wilson 1856-1924," [Online]. Available: <https://www.oxfordreference.com/display/10.1093/acref/9780191866692.001.0001/q-oro-ed6-00011630>. [Data adresu: 15 marca 2025 r.].
- [138] "Convertors", 2024. [Online]. Dostępne: <https://datadrivenconstruction.io/index.php/convertors/>.
- [139] PWC, "Sizing the prize What's the real value of AI for your business and how can you capitalise?", 1 stycznia 2017 r. [Online]. Dostępne: <https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>. [Data dostępu: 18 lutego 2025 r.].
- [140] "Rurociąg w budowie", 2024 r. [Online]. Dostępne: <https://datadrivenconstruction.io/index.php/pipeline-in-construction/>.
- [141] Wikipedia, "Apache NiFi," 1 stycznia 2025 r. [Online]. Dostępny: https://de.wikipedia.org/wiki/Apache_NiFi. [Data dostępu: 5 marca 2025 r.].
- [142] n8n, "Gmail AI Auto-Responder: Tworzenie wersji roboczych odpowiedzi na przychodzące wiadomości e-mail", 1 maja 2024 r. [Online]. Dostępne: <https://n8n.io/workflows/2271-gmail-ai-auto-responder>.

- responder-create-draft-replies-to-incoming-emails/. [Dostęp 15 marca 2025].
- [143] n8n, "Real Estate Daily Deals Automation with Zillow API, Google Sheets and Gmail", 1 marca 2025 r. [Online]. Dostępne: <https://n8n.io/workflows/3030-real-estate-daily-deals-automation-with-zillow-api-google-sheets-and-gmail/>. [Data dostępu: 15 marca 2025 r.].
- [144] B. T. O'Neill, "Wskaźniki niepowodzeń projektów analitycznych, AI i big data = 85% - yikes!", 1 stycznia 2025 r. [Online]. Dostępne: <https://designingforanalytics.com/resources/failure-rates-for-analytics-bi-iot-and-big-data-projects-85-yikes/>.
- [145] J. Neyman, On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection, Oxford University Press, 1934.
- [146] T. J. S. a. J. S. Jesse Perla, "A Problem that Stumped Milton Friedman," Quantitative Economics with Julia, 1 stycznia 2025. [Online]. Dostępne: https://julia.quantecon.org/dynamic_programming/wald_friedman.html. [Data dostępu: 1 maja 2024 r.].
- [147] T. Landsall-Welfair, Forecasting the nation's current mood, Significance, 2012.
- [148] A. Boyko, "San Francisco. Sektor budowlany 1980-2019", 2024. [Online]. Dostępne: <https://www.kaggle.com/search?q=San+Francisco.+Building+sector+1980-2019>.
- [149] A. Boyko, "Kaggle: RVT IFC Files 5000 Projects", 2024. [Online]. Dostępne: <https://www.kaggle.com/datasets/artemboiko/rvtifc-projects>.
- [150] CFMA, "Preparing for the Future with Connected Construction," [Online]. Dostępne: <https://cfma.org/articles/preparing-for-the-future-with-connected-construction>. [Dostęp 15 marca 2025 r.].
- [151] Cisco, "Cisco Survey Reveals Close to Three-Fourths of IoT Projects Are Failing", 22 maja 2017 r. [Online]. Dostępne: <https://newsroom.cisco.com/c/r/newsroom/en/us/a/y2017/m05/cisco-survey-reveals-close-to-three-fourths-of-iot-projects-are-failing.html>.
- [152] "Warunki wymagane do zachowania skamieniałości roślin", 2024. [Online]. Dostępne: <https://ucmp.berkeley.edu/IB181/VPL/Pres/PresTitle.html>.
- [153] "Fink z BlackRock o obligacjach, fuzjach i przejęciach, recesji w USA i wyborach: pełny wywiad", 2023. [Online]. Dostępne: <https://www.bloomberg.com/news/videos/2023-09-29/blackrock-s-fink-on-m-a-recession-election-full-intv-video>.
- [154] cio, "12 słynnych katastrof AI", 02 października 2024 r. [Online]. Dostępne: <https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html>. [Data dostępu: 15 marca 2025 r.].

- [155] G. Kasparow, *Głębokie myślenie*, PublicAffairs, 2017.
- [156] Wikipedia, "Kaggle", 1 stycznia 2025 r. [Online]. Dostępny: <https://en.wikipedia.org/wiki/Kaggle>. [Data dostępu: 15 marca 2025 r.].
- [157] Kaggle, "Titanic - Machine Learning from Disaster", 1 stycznia 2025 r. [Online]. Dostępne: <https://www.kaggle.com/competitions/titanic/overview>. [Dostęp 10 marca 2025 r.].
- [158] Ш. Johri, "Tworzenie ChatGPT: od danych do dialogu", 2024. [Online]. Dostępne: <https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/>.
- [159] П. Domingos, "Kilka przydatnych rzeczy, które warto wiedzieć o uczeniu maszynowym", 2024. [Online]. Dostępne: <https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf>.
- [160] J. Saramago, "Cytat", [Online]. Dostępny: <https://www.goodreads.com/quotes/215253-chaos-is-merely-order-waiting-to-be-deciphered>. [Data dostępu: 17 marca 2025].
- [161] NVIDIA, "Enhance Your Training Data with New NVIDIA NeMo Curator Classifier Models", 19 grudnia 2024 r. [Online]. Dostępne: <https://developer.nvidia.com/blog/enhance-your-training-data-with-new-nvidia-nemo-curator-classifier-models/>. [Dostęp 25 marca 2025 r.].
- [162] "NVIDIA ogłasza główne wydanie modeli Cosmos World Foundation i narzędzi danych fizycznej sztucznej inteligencji", 18 marca 2025 r. [Online]. Dostępne: <https://nvidianews.nvidia.com/news/nvidia-announces-major-release-of-cosmos-world-foundation-models-and-physical-ai-data-tools>. [Data dostępu: 25 marca 2025 r.].
- [163] NVIDIA, "NVIDIA Isaac Sim", [w Internecie]. Dostępny: <https://developer.nvidia.com/isaac/sim>. [Dostęp 25 marca 2025 r.].
- [164] M. Quarterly, "Dlaczego strategie cyfrowe zawodzą", 25 stycznia 2018 r. [Online]. Dostępne: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-digital-strategies-fail>. [Data dostępu: 15 marca 2025 r.].
- [165] M. J. Perry, "Moje ulubione cytaty Miltona Friedmana", 17 listopada 2006 r. [Online]. Dostępny: <https://www.aei.org/carpe-diem/my-favorite-milton-friedman-quotes/>. [Data dostępu: 1 marca 2025].
- [166] J. A. Wheeler, "Informacja, fizyka, kwanty: poszukiwanie powiązań", 1990.
- [167] A. Boyko, "Wojny lobbingowe i rozwój BIM. Część 5: BlackRock jest mistrzem wszystkich technologii. Jak korporacje kontrolują otwarty kod źródłowy", 2024. [Online]. Dostępne: <https://boikoartem.medium.com/lobbyist-wars-and-the-development-of-bim-d72ad0111a7d>.
- [170] T. Krijnen i J. Beetz, "A SPARQL query engine for binary-formatted IFC building models", *Advanced Engineering Informatics*, 2024.

- [171] "Liczba firm w sektorze budowlanym w Wielkiej Brytanii w 2021 r. według wielkości firmy", 2024. [Online]. Dostępne: <https://www.statista.com/statistics/677151/uk-construction-businesses-by-size/>.
- [172] "5 000 projektów IFC&RVT", 2024. [Online]. Dostępne: <https://www.kaggle.com/code/artemboiko/5000-projects-ifc-rvt-datadrivenconstruction-io>.
- [173] M. Popova, "It from Bit: Pioneering Physicist John Archibald Wheeler on Information, the Nature of Reality, and Why We Live in a Participatory Universe", 2008. [Online]. Dostępne: <https://www.themarginalian.org/2016/09/02/it-from-bit-wheeler/>. [Data odniesienia: luty 2025].
- [174] *Wojny lobbingowe o dane w budownictwie / Technofeudalizm i historia ukrytej przeszłości BIM*. [Film]. Niemcy: Artyom Boiko, 2023.
- [175] A. Boyko, "CHATGPT WITH REVIT AND IFC | Automatyczne pobieranie dokumentów i danych z projektów", 16 listopada 2023. [Online]. Dostępne: https://www.youtube.com/watch?v=ASXolti_YPs&t. [Data dostępu: 2 marca 2025 r.].
- [176] M. & Company, "Three new mandates for capturing a digital transformation's full value", 22 stycznia 2022 r. [Online]. Dostępne: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/three-new-mandates-for-capturing-a-digital-transformations-full-value>. [Data dostępu: 15 lutego 2025 r.].
- [177] KPMG, "Budownictwo w cyfrowym świecie", 1 maja 2021 r. [Online]. Dostępne: <https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf>. [Data dostępu: 5 kwietnia 2024 r.].
- [178] LLP, KPMG, "Cue Construction 4.0: Make-or-Break Time", 17 marca 2023 r. [Online]. Dostępne: <https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [Data dostępu: 15 lutego 2025 r.].
- [179] O. Business, "Satya Nadella Reveals 'How AI Agents Will Disrupt SaaS Models'", 10 stycznia 2025 r. [Online]. Dostępne: <https://www.outlookbusiness.com/artificial-intelligence/microsoft-ceo-satya-nadella-reveals-how-ai-agents-will-disrupt-saas-models>. [Data dostępu: 15 marca 2025 r.].
- [180] Forbes, "Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says," 23 Mar. 2016. [Online]. Dostępne: <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/>. [Data dostępu: 15 marca 2025 r.].
- [181] Biuro Spraw Zagranicznych, Wspólnoty Narodów i Rozwoju Wielkiej Brytanii, "Strategia rozwoju cyfrowego na lata 2024-2030", 18 marca 2024 r. [Online]. Dostępne: <https://www.gov.uk/government/publications/digital-development-strategy-2024-to-2030/digital-development-strategy-2024-to-2030>. [Data dostępu: 15 marca 2025 r.].

- [182] "Wizja i strategia w branży projektowania budynków", 7 listopada 2003 r. [Online]. Dostępne: <https://web.archive.org/web/20030711125527/http://usa.adsk.com/adsk/servlet/item?id=2255342&siteID=123112>. [Data adresu: 5 marca 2025 r.].
- [183] M. Bocharov, "Modelowanie informacji", marzec 2025. [Online]. Dostępne: <https://www.litres.ru/book/mihail-evgenievich-bocharov/informacionnoe-modelirovaniye-v-rossii-71780080/chitat-onlayn/?page=5>. [Data adresu: 15 marca 2025 r.].
- [184] "Integrated Design-Through-Manufacturing: Benefits and Rationale", 2000. [Online]. Dostępne: https://web.archive.org/web/20010615093351/http://www3.autodesk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Data odniesienia: 25 marca 2025 r.].
- [185] CAD Vendor, "Program Open BIM to kampania marketingowa", 12 marca 2012. [Online]. Dostępne: <https://web.archive.org/web/20120827193840/http://www.graphisoft.com/openbim/>. [Data odniesienia: 30 marca 2025 r.].

INDEKS PRZEDMIOTÓW

- 3D, 8, 14, 71, 73, 84, 191, 210, 215, 232, 234, 263, 276, 277, 287, 298, 299, 302, 303, 306, 307, 337, 375, 393, 394, 448, 468, 480
- 4D, 84, 172, 196, 199, 210, 229, 234, 237, 287
4IR, 43
- 5D, 84, 172, 196, 210, 237, 287, 492
- 6D, 172, 196, 229, 232, 233, 234, 235
- 7D, 232, 233, 234, 287
- 8D, 172, 196, 229, 232, 233, 234, 287
- A**
- AI, 3, 50, 52, 100, 102, 103, 106, 107, 116, 395, 457, 459, 461, 477
AIA, 289
AIM, 3, 289
AMS, 14, 84, 151, 153
Apache Airflow, 311, 361, 362, 363, 364, 366, 367, 369, 371, 399, 494
Apache NiFi, 116, 311, 361, 367, 368, 371, 399, 482, 494
Apache ORC, 62, 330, 378
Apache Parquet, 62, 67, 380, 381, 494
API, 54, 94, 95, 97, 109, 137, 138, 154, 168, 218, 219, 256, 257, 260, 271, 273, 294, 295, 296, 297, 300, 308, 326, 329, 342, 366, 369, 370, 488
- B**
- BDS, 258, 259
Big Data, 9, 67, 245, 383
BIM, 2, 3, 4, 6, 3, 14, 17, 24, 56, 58, 60, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 84, 92, 137, 138, 139, 140, 141, 142, 144, 150, 154, 156, 166, 172, 183, 184, 186, 187, 190, 195, 196, 208, 210, 211, 213, 216, 217, 218, 219, 220, 221, 222, 227, 228, 237, 238, 239, 242, 243, 246, 250, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 266, 271, 272, 273, 275, 276, 277, 278, 279, 280, 282, 285, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 308, 309, 324, 328, 337, 344, 351, 356, 358, 361, 367, 372, 376, 383, 389, 402, 403, 405, 413, 416, 431, 448, 462, 466, 475, 476, 480, 492, 494, 497
CAE, 16, 283
CAFM, 14, 24, 62, 84, 151, 153, 172, 233, 278, 326, 356, 387, 458, 473
CAM, 16, 78
CAPEX, 14, 82
CDE, 84, 175, 388, 389, 390
ChatGPT, 103, 104, 109, 110, 124, 129, 221, 303, 307, 343, 349, 421
Claude, 103, 104, 109, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443
CO₂, 72
CO₂, 229, 234, 235, 236, 237, 238, 239, 243
COBie, 156, 289, 292.
CoE, 56, 168, 169, 170, 477
Copilot, 114, 116, 459
CPIXML, 143, 272, 273, 276, 277, 279, 285, 296
CPM, 14, 17, 62, 166, 175, 233, 326, 473
CQMS, 14, 84, 177, 178, 462
CRM, 109, 369, 459.
CRUD, 51, 459
CSG, 263
CSV, 61, 62, 88, 89, 120, 123, 128, 129, 130, 131, 135, 168, 268, 272, 280, 329, 333, 344, 354, 356, 373, 377, 378, 379, 380, 401, 407, 414, 474, 480

D

DAE, 276, 277, 278, 280, 281, 284, 285, 414
 DAG, 362, 363, 365, 366
 Dash, 320, 336, 337.
 Zarządzanie danymi, 373, 395, 396, 398, 400, 401
 Data Lake, 214, 373, 376, 384, 385, 386, 387, 388, 389, 390, 400
 Data Lakehouse, 67, 373, 386, 387, 388
 Minimalizm danych, 373, 395, 396, 400, 401
 Data Swamp, 373, 395, 397, 401
 Hurtownia danych, 382, 383, 400
 Data-as-a-Service, 487
 oparte na danych, 50, 170, 460, 461, 484, 486
 DataFrame, 67, 117, 121, 122, 123, 125, 129, 130, 131, 133, 134, 135, 137, 219, 220, 224, 237, 308, 328, 329, 330, 332, 333, 344, 345, 347, 348, 354, 365, 377, 381, 409, 414, 415, 475
 DataOps, 170, 373, 398, 399, 400, 401, 475
 DeepSeek, 103, 104, 107, 109, 110, 124, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 421, 429, 443, 494
 DGN, 8, 140, 186, 227, 357
 DWG, 8, 70, 71, 73, 97, 140, 186, 211, 227, 272, 287, 302, 303, 304, 307, 357, 376, 497
 DWH, 67, 373, 376, 382, 383, 384, 386, 387, 388, 389, 390
 DXF, 8, 73, 277

E

ECM, 58, 175
 ECS, 142
 EIR, 289
 eLOD, 289
 ELT, 384, 385
 EPM, 14, 166, 198
 ERP, 2, 11, 12, 14, 17, 18, 24, 25, 58, 62, 109, 153, 166, 172, 175, 196, 198, 210, 232, 239, 240, 241, 242, 243, 244, 245, 246, 247, 249, 272, 277, 278, 279, 282, 326, 351, 356, 361, 369, 387, 388, 389, 390, 458, 462, 473, 475, 487, 492
 ESG, 196, 235, 236, 238
 ETL, 6, 1.1-8, 32, 49, 81, 113, 116, 119, 128, 188, 193, 219, 291, 311, 312, 317, 323, 324, 325, 326, 327, 328, 330, 331, 333, 338, 339, 340, 343, 344, 348, 349, 350, 351, 353, 354, 356, 361, 362, 363, 364, 365, 367, 371, 372, 381, 382, 383, 384, 385, 399, 409, 430, 475, 481, 494
 Excel, 57, 61, 62, 65, 66, 85, 88, 111, 120, 123, 123, 125, 132, 154, 167, 187, 210, 223, 224, 226, 227, 228, 278, 291, 329, 333, 342, 351, 365, 376, 378, 459, 475, 497
 Wyciąg, 81, 128, 134, 193, 311, 323, 324, 326, 328, 330, 345, 361, 362, 363, 365, 383, 384, 450, 475

F

Pióro, 62, 123, 330, 378
 FPDF, 339, 340, 341, 342, 343

G

RODO, 109
 GIS, 58
 GLTF, 143, 278
 Arkusze Google, 368, 370
 Grok, 103, 104, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443

H

HDF5, 62, 67, 123, 329, 330, 378, 379, 380
 HiPPO, 29, 37, 95, 424, 477, 484, 490
 HTML, 123, 340, 365, 370

I

IDS, 289, 290, 291.
 IFC, 8, 73, 138, 142, 186, 227, 261, 262, 263, 264, 265, 266, 267, 268, 272, 273, 273, 276, 277, 278, 279, 280, 284, 286, 292, 296, 302, 329, 357, 414, 415, 417, 497
 IGES, 262, 263, 276.
 iLOD, 289
 IoT, 10, 18, 67, 271, 367, 369, 405, 413, 417, 417, 418, 419, 455, 460, 465, 482, 484
 ISO 19650, 388

J

JavaScript, 320, 369, 378
 JSON, 88, 89, 90, 92, 123, 128, 142, 269, 272, 280, 329, 330, 333, 378, 474, 480
 Jupyter Notebook, 114, 115, 116, 130, 187, 224, 330, 346, 417, 425

K

Kaggle, 115, 121, 130, 187, 224, 303, 307, 330, 346, 408, 415, 417, 425, 426, 430, 431, 433
 k-NN, 392, 393, 442, 445, 446, 447, 448
 KPI, 245, 311, 317, 318, 319, 320, 321, 324, 353, 372, 389, 478, 479

L

LEED, 235, 236, 238.
 LlaMa, 103, 104, 120, 124, 129, 133, 166, 219, 300, 329,

343, 349, 381, 421, 429, 443, 494

LLM, 3, 4, 24, 29, 50, 51, 52, 55, 56, 92, 95, 99, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 118, 120, 124, 125, 126, 129, 131, 133, 134, 135, 166, 166, 187, 219, 220, 220, 221, 222, 222, 224, 225, 226, 231, 237, 238, 251, 294, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 328, 329, 332, 333, 334, 335, 336, 338, 339, 340, 343, 344, 345, 346, 347, 348, 349, 354, 370, 372, 376, 381, 382, 392, 394, 399, 413, 425, 429, 430, 432, 433, 435, 443, 444, 457, 460, 461, 476, 488, 494

Obciążenie, 81, 128, 134, 193, 311, 323, 325, 326, 333, 334, 338, 339, 342, 343, 361, 362, 363, 365, 367, 383, 384, 450, 475

LOD, 287, 289

LOI, 287

LOMD, 287

Low-Code, 368, 369

M

Matplotlib, 123, 306, 320, 333, 335, 342, 372, 409, 411, 494

MCAD, 77, 78, 257, 284

MEP, 14, 175, 492

MESH, 234, 283, 284, 285, 296, 462

Microsoft SQL, 65

Mistral, 103, 104, 107, 110, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 421, 429, 443, 494

MRP, 11, 12

MS Project, 70

MySQL, 63, 64, 65, 291, 329, 330

N

n8n, 116, 311, 361, 368, 369, 370, 371, 372

NLP, 69, 394

No-Code, 368, 369

NURBS, 142, 282, 283, 284, 285

NWC, 8, 276

O

OBJ, 143, 273, 276, 277, 278, 280, 281, 284, 285

OCCT, 273

OCR, 69, 128, 132, 134

OmniClass, 154, 155, 156.

otwarty BIM, 142, 216, 219, 256, 261, 278, 291

Open Source, 45, 55, 97, 98, 107, 108, 273, 275, 494

SOWA, 267, 268, 269.

P

Pandy, 56, 67, 103, 117, 118, 119, 120, 121, 122, 123, 125, 130, 134, 138, 186, 187, 220, 221, 225, 227, 269, 300, 303, 328, 329, 330, 377, 378, 380, 381, 403, 408, 409, 410, 411, 412, 414, 475, 479, 494

Parkiet, 67, 123, 329, 330, 373, 378, 379, 380, 381, 382, 386, 401, 403, 414, 474, 480

PDF, 69, 70, 71, 85, 111, 126, 127, 128, 129, 130, 131, 132, 134, 146, 147, 177, 186, 190, 191, 211, 215, 278, 338, 339, 340, 341, 342, 342, 344, 345, 346, 347, 354, 355, 356, 357, 365, 376, 475, 480, 497

PDM, 16

PHP, 63, 339

Rurociąg, 44, 53, 115, 128, 183, 303, 307, 308, 311, 312, 349, 350, 351, 352, 354, 355, 356, 367, 370, 417, 430, 476

PLM, 16, 246

ZŁ, 8, 272, 296

Plotly, 320, 336, 337, 372

PMIS, 3, 24, 32, 62, 152, 196, 198, 211, 239, 240, 245, 246, 247, 248, 249, 250, 272, 326, 351, 387, 388, 389, 390, 458, 462, 487

PMS, 84, 151, 233

PostgreSQL, 63, 64, 65, 329, 395

Power BI, 320, 372, 482

private equity, 464, 487

Python, 56, 63, 103, 105, 112, 114, 115, 117, 118, 119, 129, 130, 131, 133, 134, 135, 166, 167, 179, 187, 219, 224, 225, 227, 303, 304, 320, 329, 330, 332, 333, 339, 340, 346, 348, 356, 367, 369, 372, 378, 409, 417, 426, 459, 475, 476, 479, 494

Q

QTO, 72, 196, 214, 215, 216, 217, 218, 219, 221, 223, 225, 226, 228, 237, 238, 242, 243, 251, 301, 302, 475, 479

QWEN, 103, 104, 124, 129, 133, 166, 219, 221, 300, 329, 349, 343, 349, 381, 429, 443

R

RAG, 111, 116

RDBMS, 63, 64, 65, 82, 89.

RDF, 267, 268, 269

RegEx, 126, 136, 177, 179, 331, 332, 333, 357, 474

RFID, 8, 18, 58, 84, 418, 419, 455, 460, 465, 482

ROI, 311, 317, 319, 321, 351, 370, 389, 479

RPM, 14, 84, 151, 331, 333, 334

RTV, 8, 73, 77, 140, 186, 227, 272, 296, 300, 302, 322, 329, 357, 376, 414, 415, 417, 497

S

SaaS, 24, 50, 51, 52, 458
ZAKRES, 273, 277
SDK, 139, 141, 257, 264, 273, 281, 286, 296, 329, 414
Seaborn, 123, 320, 336, 337, 408, 412
SPARQL, 269
SQL, 63, 65, 66, 88, 89, 103, 105, 119, 123, 166, 168, 268, 269, 279, 276, 277, 300, 329, 376, 392, 494
SQLite, 63, 64, 65, 166, 167, 296, 329, 363
KROK, 261, 262, 263, 266, 272, 276, 277, 292
Streamlit, 336, 337
SVF, 142, 276, 285

T

Transform, 128, 193, 311, 323, 325, 326, 330, 331, 333, 334, 340, 341, 342, 344, 347, 361, 362, 363, 365, 367, 383, 384, 450, 475

U

Uniclass, 154, 155, 156.
USD, 142, 143, 255, 276, 277, 278, 279, 280, 281, 284, 285, 286, 292, 296

V

VectorOps, 373, 398, 400, 401
VR, 84, 271, 285

W

WhiteBox, 240, 242, 243.

X

XLSX, 8, 61, 62, 123, 128, 129, 231, 268, 276, 277, 280, 296, 299, 308, 351, 373, 377, 378, 379, 380, 401, 474, 480
XML, 61, 88, 89, 92, 128, 143, 269, 272, 277, 280, 280, 291, 299, 329, 378, 414, 474